



Filipa Isabel Figueira Simão

Licenciada em Ciências Forenses e Criminais

Estudo do ADN mitocondrial de uma população de indivíduos nativos de Angola residentes em Lisboa

Dissertação para obtenção do Grau de Mestre em
Bioquímica

Orientador: António Amorim, Especialista Superior de Medicina Legal, Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação Sul

Co-orientadora: Heloísa Afonso Costa, Especialista Superior de Medicina Legal, Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação Sul

Júri:

Presidente: Prof. Doutor Pedro António de Brito Tavares

Arguente: Prof. Doutora Maria Marise Simões de Almeida

Vogal: Mestre Heloísa Afonso Costa



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Dezembro 2014

Filipa Isabel Figueira Simão

Licenciada em Ciências Forenses e Criminais

Estudo do ADN mitocondrial de uma população de indivíduos nativos de Angola residentes em Lisboa

Dissertação para obtenção do Grau de Mestre em
Bioquímica

Orientador: António Amorim, Especialista Superior de Medicina Legal, Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação Sul

Co-orientadora: Heloísa Afonso Costa, Especialista Superior de Medicina Legal, Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação Sul

Monte da Caparica, Faculdade de Ciências e Tecnologia, Universidade de Lisboa

Dezembro 2014

Estudo do ADN mitocondrial de uma população de indivíduos nativos de Angola residentes em Lisboa

Copyright ©

Eu, Filipa Isabel Figueira Simão, declaro que a Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

Agradecimentos

No final desta etapa não podia deixar de agradecer a quem contribuiu, a nível profissional e pessoal, para a sua concretização.

Ao meu orientador Professor António Amorim, por ter aceite a orientação da presente Dissertação de Mestrado, pela sua disponibilidade e pelo grande apoio prestado na realização deste trabalho.

À minha orientadora Mestre Heloísa Afonso Costa, por ter aceite a orientação da presente Dissertação de Mestrado, por todos os conhecimentos transmitidos, pela tranquilidade demonstrada e pelos incentivos.

Ao Sr. Prof. Jorge Costa Santos e ao Dr. Mário Dias, na condição de anterior e atual diretor da delegação sul do Instituto Nacional de Medicina Legal, por permitirem a realização do projeto nas instalações do serviço de Genética e Biologia Forenses do Instituto Nacional de Medicina Legal e Ciências Forenses.

À Dra. Teresa Ribeiro por me ter recebido e integrado no Serviço de Genética e Biologia Forenses.

À restante equipa do Serviço de Genética e Biologia Forenses, pela ajuda técnica prestada.

À Rita Reis e à Joana Lopes pela amizade que surgiu e cresceu para além das paredes do laboratório. À Carina Almeida, Catarina Dourado e restantes colegas estagiárias pela boa disposição e momentos de descontração.

À Margarida Cassiano, Filipa Loureiro, Emma Gomes, Maria Carrega, Raquel Cláudio, Luís Salvador, Luís Guerreiro, João Paulos e Diogo Lança pela amizade. Cada um, da sua forma e no seu momento, me incentivou e apoiou. Estiveram presentes nos momentos cruciais.

Aos meus pais, Maria João Simão e José Simão, por sempre confiarem e apoiarem as minhas escolhas e decisões e por me demonstrarem que com trabalho e dedicação é possível ultrapassar todos os obstáculos. Ao meu irmão Miguel Simão, pela facilidade com que me faz sorrir, mesmo nos dias menos alegres.

Resumo

Nas últimas décadas o número de imigrantes em Portugal aumentou consideravelmente. Até ao final do ano de 2013 o número de imigrantes Angolanos a residir em Portugal atingiu os 20 000 indivíduos, verificando-se uma predominância territorial na região metropolitana de Lisboa.

Angola é um país situado na costa Atlântica Sul do continente Africano. A presença do povo Bantu, assim como a colonização pelo povo Português durante o século XV, são considerados como os principais modeladores do padrão genético da população deste país.

A transmissão exclusiva por via materna, o elevado número de cópias, a ausência de recombinação e a elevada taxa de mutação inerentes ao ADN mitocondrial, são características que o tornam útil em estudos de origem e evolução Humana, assim como em investigações forenses.

O principal objetivo deste estudo foi caracterizar geneticamente a população de imigrantes Angolanos a residir em Lisboa. Para tal foi estudado um grupo de 173 indivíduos, residentes em Lisboa, com ascendência Angolana confirmada e não relacionados entre si. Sequenciou-se a região controlo do ADN mitocondrial com recurso aos *primers* L15997/H016 e L16555/H639.

Cerca de 85% dos haplótipos identificados são únicos. A maioria dos haplogrupos determinados pertence a linhagens de ADN mitocondrial descritas como específicas da região subsariana de África, com cerca de 87% dos haplótipos pertencentes ao macrohaplogrupo L.

Do estudo filogenético verificou-se que as populações geneticamente mais próximas foram a nossa população imigrante de Angola e as populações de indivíduos Angolanos a residir em Angola e de indivíduos pertencentes a diversos grupos étnico-linguísticos Bantu.

Este estudo vem alertar para a grande diversidade genética que a população imigrante Angolana introduz em Lisboa nas gerações atuais e nas gerações futuras. Teremos num futuro muito próximo, indivíduos naturais e nacionais de Lisboa com haplótipos, até então, considerados como tipicamente Africanos.

Os haplótipos da população Angolana imigrante a residir em Lisboa foram submetidos e aceites para inserção na base de dados EMPOP (*EDNAP Forensic mtDNA Population Database*) com o número de acesso EMPOP662.

Palavras-chave: ADN mitocondrial; região controlo total; Angola; Lisboa

Abstract

In the past decades the number of immigrant people in Portugal has grown. According to statistical data, until the end of 2013 the total number of Angolan immigrants in Portugal equalled about 20 000 individuals. A territorial predominance has been found for the metropolitan region of Lisboa. Angola is a country located at Atlantic coast of Africa. The presence of Bantu people and the colonization by Portuguese people on Angolan territory are considered to be the major modulators of genetic patterns in Angola.

Maternal inheritance, high copy number, lack of recombination and high mutation rate inherent to mitochondrial DNA, are characteristics that makes this genetic marker useful in studies of human origin and evolution. Mitochondrial DNA can also contribute to ascertain the identity of individuals in forensic cases, by determining the haplotypes frequency.

The main aim of this study was to ascertain the genetic structure of Angolan immigrant population living in Lisboa. Therefore a total of 173 individuals, inhabitants at Lisboa, nonrelated and with Angolan ancestry, were studied. Total control region of mitochondrial DNA was amplified using the *primers* L15997/H016 and L16555/H639.

About 85% of the studied haplotypes are unique among the 173 samples. The majority of the identified haplotypes belong to mtDNA lineages known to be specific of sub-Saharan region (about 87% haplotypes belong to macrohaplogroup L).

Through the phylogenetic study it was found that the populations genetically closer were our immigrant population of Angola and the populations of individuals from various ethnic and linguistic Bantu groups.

This study evidences the genetic diversity that Angolan immigrant population introduces in Lisbon, on current and future generations. We will have, in a very near future, natural and national individuals living in Lisbon with haplotypes, until recently, considered as typically African.

Lisboa immigrant population from Angola can be accessed via EMPOP (EDNAP Forensic mtDNA Population Database) dataset with accession number EMPOP662.

Keywords: Mitochondrial DNA; control region; Angola; Lisboa

Índice

Agradecimentos.....	i
Resumo.....	iii
<i>Abstract</i>	v
Índice.....	vii
Índice de figuras.....	ix
Índice de tabelas.....	xi
Abreviaturas e siglas.....	xiii
1. Objetivos	1
2. Introdução.....	3
2.1 A origem do Homem moderno	3
2.2 Angola.....	4
2.2.1 Aspetos geográficos.....	4
2.2.2 História de Angola.....	5
a. Período pré-colonial.....	6
b. Período colonial	7
c. Angola de hoje	7
2.3 O fenómeno de Migração em Portugal.....	9
2.4 Mitocôndria	10
2.4.1 Estrutura da mitocôndria	10
2.4.2 Função da mitocôndria.....	10
2.4.3 Origem da mitocôndria e Hipótese endossimbiótica.....	10
2.5 ADN mitocondrial.....	12
2.5.1 Estrutura do ADNmt	14
2.5.2 Características do ADNmt.....	14
a. Elevado número de cópias	14
b. Herança materna	15
c. Ausência de recombinação.....	16
d. Elevada taxa de mutação	16

2.5.3 Sequência de Referência de Cambridge - CRS	18
2.5.4 Eva mitocondrial.....	19
2.5.5 Heteroplasmia	19
2.5.6 Aplicações do genoma mitocondrial	21
a. ADNmt e genética populacional.....	21
b. Aplicação forense do ADNmt.....	23
2.5.7 Bases de dados de ADNmt.....	24
3. Procedimento experimental.....	25
3.1 Amostragem.....	26
3.2 Extração de ADN.....	26
3.3 Amplificação.....	27
3.4 Purificação dos produtos amplificados	29
3.5 Sequenciação.....	30
3.6 Purificação dos produtos sequenciados	32
3.7 Detecção dos produtos sequenciados.....	32
3.8 Análise dos produtos sequenciados.....	33
3.9 Análise estatística e filogenética	37
4. Resultados	39
4.1 Composição nucleotídica.....	42
4.2 Heteroplasmia	46
4.3 Haplogrupos.....	47
4.4 Cálculos da diversidade	48
4.5 Análise filogenética.....	51
5. Discussão	57
5.1 Região controlo total do ADNmt.....	58
5.2 Discussão dos resultados obtidos	60
5.2.1 Composição nucleotídica	61
5.2.2 Haplótipos	62
5.2.3 Heteroplasmia	63
5.2.4 Haplogrupos	64
5.2.5 Parâmetros de diversidade	67
5.2.6 Significado filogenético	68
6. Conclusões	71
7. Bibliografia	73
8. Anexos.....	83

Índice de figuras




Figura 1 – Padrão de dispersão humana durante os 100 000 anos passados (Henn <i>et al.</i> 2012). O símbolo  representa o <i>founder effect</i> ; o símbolo  indica a fonte do <i>founder effect</i> ; e o símbolo  assinala a direção do movimento migratório. Kya vem do inglês <i>thousand years ago</i> , significa mil anos atrás.	3
Figura 2 – Distribuição das 18 províncias de Angola (http://www.diarioliberdade.org/archivos/Colaboradores_avanzados/manu/2013-08/Foto_-_Wikicommons.png)	5
Figura 3 – Migração do povo Bantu (adaptado de http://www.face-music.ch/artuganda/bantumigration.jpg)	6
Figura 4 – Distribuição das tribos Bantu no território Angolano no ano de 1970 (http://upload.wikimedia.org/wikipedia/commons/thumb/7/70/Angola_Ethnic_map_1970-pt.svg/488px-Angola_Ethnic_map_1970-pt.svg.png)	8
Figura 5 - Percentagem da população imigrante em Portugal no ano de 2012 (Serviço de Estrangeiros e Fronteiras 2013)	9
Figura 6 – Estrutura do ADNmt. a) Molécula de ADNmt; b) Região não codificante do ADNmt (adaptado de Butler 2011)	13
Figura 7 – Representação da transmissão materna do ADNmt referente a 18 indivíduos (Butler 2005). Os círculos representam indivíduos do sexo feminino e os quadrados indivíduos do sexo masculino.	15
Figura 8 – Mapa representativo da dispersão do Homem por todo o mundo e distribuição geográfica dos haplogrupos de ADNmt (Adaptado de http://www.familytreedna.com)	22
Figura 9 – Fluxograma representativo do procedimento experimental realizado para obtenção dos haplótipos de ADNmt	25
Figura 10 – Esquema do processo de extração de ADN através do método de Chelex® 100. (Adaptado de Butler 2005).	27
Figura 11– Representação das gamas de temperatura durante a <i>polymerase chain reaction</i> e intervalos de tempo correspondente (adaptado de Butler 2005)	29
Figura 12 – Sequenciação de ADN com ddNTPs marcados com corante fluorescente (adaptado de Butler 2005)	30
Figura 13 - Representação das gamas de temperatura durante a sequenciação, e intervalo de tempo correspondente (adaptado de Butler 2005).	31
Figura 14 – Sequenciador Genetic Analyser 3130 (AB) do serviço de Genética e Biologia Forenses do INMLCF-DS	33

Figura 15– Genetic Analyser 3130 (AB), vista interior. (1) Reservatório ânodo; (2) Reservatório cátodo.....	33
Figura 16 – Segmento de uma sequência de ADNmt obtido pelo <i>primer</i> L16555, analisada pelo <i>software Sequencing Analysis</i> v.5.2.	34
Figura 17 - Eletroferograma obtido pelo <i>primer</i> L15997. (1) Nucleótido que origina muitas dúvidas; (2) Nucleótido suscetível a dúvidas; (3) Nucleótidos corretamente identificado.	34
Figura 18 – Eletroferogramas de uma amostra com heteroplasma de comprimento numa cadeia de ADNmt da região HVII, analisado pelo <i>software Sequencing Analysis</i> v.5.2. a) sequenciação com o <i>primer</i> associado à cadeia leve - L16555; b) sequenciação com o <i>primer</i> associado à cadeia pesada - H639.	36
Figura 19– Heteroplasma de posição identificada a vermelho.	36
Figura 20 – Árvore filogenética de ADNmt (van Oven & Kayser 2009).....	37
Figura 21 – Percentagem dos diferentes haplogrupos identificados nas 173 sequências de ADNmt pertencentes a imigrantes Angolanos a residir em Lisboa.....	48
Figura 22 – Representação filogenética de distâncias genéticas de populações Africanas relativas à região controlo do ADNmt.	55
Figura 23 - Representação filogenética de distâncias genéticas de populações Africanas relativas aos segmentos HVI e HVII da região controlo do ADNmt.....	56
Figura 24 – <i>Network</i> obtido através do <i>software DrawNetWork</i> v1.24. com recurso ao filtro EMPOPall_R11.	62

Índice de tabelas

Tabela 1 – Comparação entre o ADN nuclear e o ADN mitocondrial (adaptado de Butler 2011).	12
Tabela 2 - Sequência dos <i>primers</i> utilizados na amplificação.	28
Tabela 3 - Volumes utilizados na preparação da mistura de reagentes necessários para amplificação da região controlo total do ADNmt.	29
Tabela 4 - Volumes utilizados na preparação da mistura de reagentes necessários para sequenciação da região controlo total do ADNmt.	31
Tabela 5 – Volumes por reação necessários no processo de purificação dos produtos sequenciados.	32
Tabela 6 – Código IUPAC utilizado na análise do ADNmt.	35
Tabela 7 - Identificação do código da sequência, respetivo haplogrupo e haplótipo com os respetivos polimorfismos referentes à rCRS.	40
Tabela 8 – Haplótipos de maior frequência nas 173 sequências estudadas no presente estudo	42
Tabela 9 – Distribuições dos polimorfismos observados na região controlo total do ADNmt nas 173 sequências de imigrantes Angolanos a residir em Lisboa.	42
Tabela 10 – Polimorfismos presentes nas 173 sequências de ADNmt estudadas, compreendidos entre as posições 16024 e 16569.	45
Tabela 11 – Heteroplasma de comprimento observada nas 173 sequências de ADNmt de imigrantes Angolanos a residir em Lisboa.	46
Tabela 12 – Cálculo do valor de π^2 para a totalidade da região controlo do ADNmt.	49
Tabela 13 – Cálculo do valor de χ^2 para a totalidade da região controlo do ADNmt.	49
Tabela 14 – Valores de diversidade de sequência (\hat{H}) e diversidade nucleotídica (π) obtidos para a totalidade da região controlo do ADNmt.	50
Tabela 15 - Valores de diversidade de sequência (\hat{H}) e diversidade nucleotídica (π) obtidos para os segmentos HVI e HVII da região controlo do ADNmt.	51
Tabela 16 - Valores de diversidade de sequência (\hat{H}) e diversidade nucleotídica (π) obtidos para o segmento HVI da região controlo do ADNmt.	51
Tabela 17 – Populações retiradas da literatura referentes à totalidade da região controlo do ADNmt, utilizadas para fins comparativos.	52
Tabela 18 – Valores de F_{st} e p obtidos através do programa Arlequin para a totalidade da região controlo do ADNmt.	52
Tabela 19 - Populações retiradas da literatura referentes aos segmentos HVI e HVII da região controlo do ADNmt, utilizadas para fins comparativos.	53

Tabela 20 - Valores de Fst e p obtidos através do programa Arlequin para os segmentos HVI e HVII da região controlo do ADNmt.	54
Tabela 21 – Mutações que apresentam uma baixa frequência na população de imigrantes Angolanos a residir em Lisboa.	61
Tabela 22 – Distribuição dos haplogrupos não Africanos identificados no conjunto de sequências dos imigrantes Angolanos a residir em Lisboa.	66

Abreviaturas e siglas

°C	grau Celcius
μM	Micromolar
A	Adenina
AB	<i>Applied Biosystems</i>
ADN	Ácido desoxirribonucleico
ADNmt	ADN mitocondrial
ADNn	ADN nuclear
ADP	Difosfato de Adenosina
AMOVA	Análise de variância molecular
ARN	Ácido ribonucleico
ATP	Trifosfato de Adenosina
C	Citosina
CRS	Sequência de referência de Cambridge (<i>Cambridge reference sequence</i>)
ddATP	Didesoxiadenina trifosfatada
ddCTP	Didesoxicitosina trifosfatada
ddGTP	Didesoxiguanina trifosfatada
ddNTP	Didesoxinucleótidos trifosfatados
ddTTP	Didesoxitimina trifosfatada
D-loop	Laço de deslocamento (<i>Displacement Loop</i>)
dNTP	Desoxinucleótidos trifosfatados
EDNAP	<i>European DNA Profiling</i>
EDTA	<i>Ethylenediamine tetraacetic acid</i>
EMPOP	<i>EDNAP forensic mtDNA population database</i>
FADH ₂	Dinucleótido de flavina e adenina (<i>Flavin adenine dinucleotide</i>)
FBI	<i>Federal Bureau of Investigation</i> (Agência Federal de Investigação)
Fst	<i>Fixation index</i> (Índice de fixação)
G	Guanina
H ₂ O	Água
HSP	Promotor da cadeia pesada (<i>Heavy-strand promotor</i>)
HVI	Região hipervariável I (<i>Hypervariable region I</i>)
HVII	Região hipervariável II (<i>Hypervariable region II</i>)
HVIII	Região hipervariável III (<i>Hypervariable region III</i>)
INMLCF	Instituto Nacional de Medicina Legal e Ciências Forenses
IUPAC	União internacional de química pura e aplicada (<i>International Union of Pure and Applied Chemistry</i>)
Km	Quilômetro
LSP	Promotor da cadeia leve (<i>Light-strand promotor</i>)
mL	Mililitro

MRCA	Ancestral comum mais recente (<i>Most Recent Common Ancestor</i>)
NADH	Dinucleótido de nicotinamida e adenina (<i>Nicotinamide adenine dinucleotide</i>)
O _H	Origem de replicação cadeia pesada
O _L	Origem de replicação cadeia leve
Pb	Pares de bases
PCR	Reação em cadeia de polimerase (<i>Polymerase Chain Reaction</i>)
rCRS	Sequência de referência de Cambridge revista (<i>Revised Cambridge reference sequence</i>)
rpm	Rotações por minuto
RFLP	<i>Restriction fragment length polymorphism</i>
SNP	Polimorfismo de nucleótido único (<i>Single nucleotide polymorphism</i>)
SEF	Serviço de Estrangeiros e Fronteiras
SGBF-DS	Serviço de genética e Biologia Forenses, Delegação Sul
SWGDAM	<i>Scientific Working Group for DNA Analysis Methods</i>
T	Timina
UV	Ultra violeta

1. Objetivos

O principal objetivo do presente estudo será realizar a caracterização genética da população imigrante Angolana a residir em Lisboa, através da análise da totalidade da região controlo do ADN mitocondrial. Espera-se que este grupo de indivíduos apresente um padrão genético mitocondrial característico de populações Africanas, particularmente das que se situam no Sudoeste do Continente Africano. Tentar-se-á perceber se a população imigrante Angolana é representativa da sua população de origem.

Com a determinação dos haplótipos dos indivíduos imigrantes e, consequentemente, a sua inclusão em haplogrupos, é possível obter informação relativa à origem geográfica do conjunto de indivíduos estudados. É também pretendido comparar a população em estudo com outras populações Africanas, o que permitirá esclarecer a sua proximidade genética.

Após a determinação dos haplótipos das sequências de ADN mitocondrial, é pretendido submeter as sequências à base de dados EMPOP - *EDNAP Forensic mtDNA Population Database* -, a mais importante base de dados de ADN mitocondrial a nível internacional para fins forenses e populacionais, de forma a disponibilizar os resultados obtidos a toda a comunidade científica.

Será possível avaliar o impacto deste conjunto de imigrantes Angolanos na região de Lisboa, clarificando a contribuição destes indivíduos na constituição genética da região de fixação.

2. Introdução

2.1 A origem do Homem moderno

O percurso evolutivo do Homem, que deu origem às diferenças entre indivíduos da população Humana atual, é consequência da contínua adaptação, guiada pela seleção natural.

A origem do Homem moderno terá acontecido em África há cerca de 45 a 60 milhões de anos. A partir desta origem terá ocorrido a dispersão dos indivíduos, primeiramente para as regiões de Eurásia e Oceânia e, posteriormente, para o resto do Mundo (Figura 1). O *founder effect* é definido como a perda de variabilidade genética que ocorre quando um número reduzido de indivíduos, pertencentes a uma população original com maior número de indivíduos, se estabelece como nova população (Henn *et al.* 2012; Oppenheimer 2012).

São duas as principais teorias que tentam explicar o fenómeno de evolução e dispersão da espécie humana - a teoria *Out Of Africa* e a teoria Multirregional. Ambas consideram que a linhagem humana teve origem no continente Africano (Figura 1) (Tattersall 2009; Stewart & Stringer 2012).

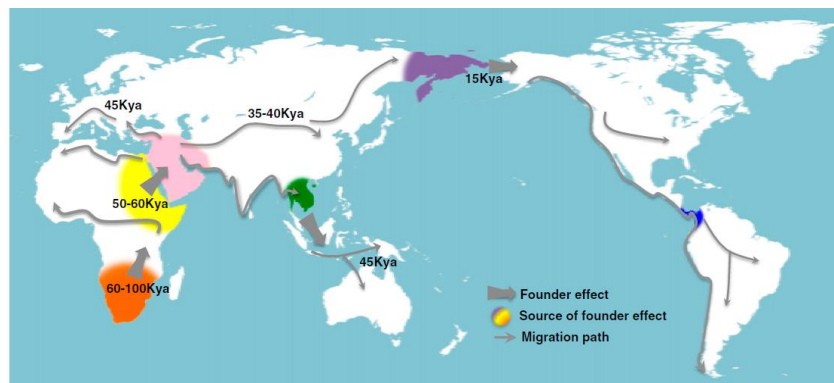





Figura 1 – Padrão de dispersão humana durante os 100 000 anos passados (Henn *et al.* 2012). O símbolo  representa o *founder effect*; o símbolo  indica a fonte do *founder effect*; e o símbolo  assinala a direção do movimento migratório. Kya vem do inglês *thousand years ago*, que significa mil anos atrás.

O modelo Multirregional (Wolpoff *et al.* 2000) considera que, após uma dispersão inicial da espécie *Homo erectus* a partir de África, a espécie *Homo sapiens* não teve origem num único local geográfico, mas surgiu de forma simultânea em diferentes locais do Mundo (Templeton 2002; Haviland *et al.* 2012; Wolpoff *et al.* 2000).

A hipótese *Out Of Africa* relativa à origem do Homem moderno surgiu nos anos 80, com a divulgação da descoberta de fósseis, a Este e Sul do continente Africano, que se assemelhavam à espécie *Homo sapiens* (Tattersall 2009). Esta teoria postula que, após a expansão inicial do *Homo erectus*, as populações de África, Europa e Ásia não mantiveram qualquer contacto e evoluíram de forma independente. O Homem moderno, terá surgido e evoluído em África, a partir de uma única população de *Homo sapiens* arcaico. Após esta fase da evolução, terá ocorrido uma nova dispersão que terá culminado na substituição e extinção global das populações humanas não modernas por populações modernas da espécie *Homo sapiens*, oriundas de África. (Templeton 2002; Tattersall 2009; Haviland *et al.* 2012; Maca-Meyer *et al.* 2001).

O pressuposto da origem do Homem moderno em África é a teoria mais aceite entre a comunidade científica. Diversos estudos realizados com recurso a marcadores genéticos autossomais e a marcadores genéticos uniparentais suportam esta hipótese (McEvoy *et al.* 2011; Laval *et al.* 2010; Gravel *et al.* 2011; Underhill & Kivisild 2007).

2.2 Angola

2.2.1 Aspetos geográficos

Angola é um país situado no hemisfério Sul, no Sudoeste do Continente Africano. Tem uma área de 1 024 670 Km² e é o segundo maior país da região subsariana de África. Faz fronteira a Norte e Este com a República Democrática do Congo, ainda a Este com a Zâmbia e a Sul com a Namíbia; a Oeste faz fronteira com o Oceano Atlântico numa extensão de 1 650 Km (Stead & Rorison 2009; Raeymaeker 2012; Zau 2002).

Angola está dividida em 18 províncias sendo elas: Bengo, Benguela, Bié, Cabinda, Cuando Cubango, Kwanza-Norte, Kwanza-Sul, Cunene, Huambo, Huíla, Luanda, Lunda-Norte, Lunda-Sul, Malange, Moxico, Namibe, Uíge e Zaire. Luanda é a capital de Angola (Figura 2).



Figura 2 – Distribuição das 18 províncias de Angola
(http://www.diarioliberalidade.org/archivos/Colaboradores_avanzados/manu/2013-08/Foto_-_Wikicommons.png, visualizado a Agosto de 2014)

A maioria do território Angolano é composto por savana (planície com árvores dispersas) mas apresenta também uma área de floresta extensa na zona de Cabinda e um excerto de deserto na fronteira da província de Namibe com a Namíbia. Os três principais rios Angolanos são o Cuanza, o Cunene e o Congo.

O país é caracterizado por duas estações climáticas distintas: uma de chuvas (húmida e quente) que ocorre entre Setembro e Abril e a estação denominada por cacimbo (seca e fria) que tem início em Maio e termina em Setembro. No entanto, sendo um território extenso, verifica-se uma grande variedade climática ao longo do país.

Os principais recursos naturais são o petróleo (sendo Angola o segundo maior produtor de petróleo na região subsariana) e os diamantes. Angola possui ainda minérios de ferro, cobre, ouro, chumbo, zinco, manganês, volfrâmio, estanho e urânio (Zau 2002; Stead & Rorison 2009; Raeymaecker 2012).

A população concentra-se maioritariamente na costa atlântica onde grande percentagem de indivíduos se dedica à atividade piscatória.

2.2.2 História de Angola

Do ponto de vista sociolinguístico a população Angolana é considerada um mosaico originado por diferentes grupos étnicos. É uma sociedade heterogénea que resultou de sucessivos movimentos migratórios humanos (Anon 1992; Anon 1975).

a. Período pré-colonial

Há cerca de 3 000 ou talvez 4 000 anos, os povos Bantu oriundos da região entre o rio Niger e o lago Tchad (Phillipson 2005; Newman 1997) começaram a deslocar-se dando origem a dois principais movimentos migratórios, para o Sul e para Este do continente Africano (Figura 3). O movimento migratório para Sul culminou com a fixação deste povo no território Angolano (Raeymaeker 2012; Redinha 2009; Anon 1992).



Figura 3 – Migração do povo Bantu (adaptado de <http://www.face-music.ch/artuganda/bantumigration.jpg>, visualizado a Agosto de 2014)

O nome Bantu tem origem no termo “ntu” utilizado por estes grupos linguístico-culturais para designar Homem, que juntamente com o prefixo “ba” refere o plural do termo; desta forma, Bantu significa Homens, seres humanos (Redinha 2009).

O movimento migratório do povo Bantu ocorreu gradualmente ao longo do tempo e levou à ocupação do território Angolano com consequente substituição e/ou expulsão do povo Khoes-san, primeiros habitantes da região. A classificação Khoes-san deriva da junção do termo Khoes, utilizado para designar indivíduos dedicados à pastorícia, e o termo San, que significa caçadores recolectores.

A fixação dos povos Bantu em diversos territórios da região Angolana com aproveitamento dos recursos naturais existentes originou diferentes reinos: Congo, Ndongo, Luanda, Bié, Bduu, Huambo, entre outros. Os indivíduos pertencentes a cada um destes Reinos desempenhavam diferentes funções e atividades, de acordo com o local em que se fixavam. Nos reinos mais nórdicos como o do Congo e de Luanda, com a proximidade dos rios Angolanos e a existência de um clima húmido, as populações a dedicavam-se à agricultura. Por outro lado, nas regiões de planalto, com a abundância de chuvas e terras férteis, os povos dedicavam-se à agricultura e à criação de gado. Na região sul, como as chuvas eram menos frequentes e irregulares e o clima mais seco, os povos viviam da pastorícia.

b. Período colonial

No século XV, com a ascensão do rei D. João II ao trono Português e o seu incentivo à descoberta do caminho marítimo para a Índia, o reconhecimento da costa Africana progrediu rapidamente. No seguimento de uma expedição naval organizada pelo rei de Portugal, o navegador Diogo Cão ancorou na foz do rio Congo no ano de 1482. A denominação de Angola ao território surgiu através do contacto dos portugueses com o soberano do Reino Ndongo, que se chamava N'gola (Figueiredo 1945; Redinha 2009).

No período inicial de ocupação de Angola, o interesse do reino de Portugal no território, para além do comércio de matérias-primas e a conversão religiosa dos nativos, passava pela comercialização de escravos. Durante o século XV foram também descobertos, pelos navegadores Espanhóis e Portugueses, territórios no continente americano, para onde eram transportados escravos Africanos, para executar o trabalho nas plantações e minas (Anon 1975; Anon 1992; Thomas 1999).

O desenvolvimento de Luanda atraiu o interesse de países europeus como França, Espanha e Holanda, que também pretendiam obter matérias-primas e escravos de África. Em 1641 os holandeses, que navegavam a costa Angolana desde o início do século XVII, apoderaram-se de Luanda, Benguela e outras regiões litorais do país. A invasão, que teve como principal objetivo a comercialização de escravos e seu transporte para o Nordeste do Brasil, durou até 1648, ano em que Portugal libertou o território por ordem do rei D. João IV (Stead & Rorison 2009; Figueiredo 1945).

Foi apenas no século XIX que os portugueses começaram uma exploração mais intensiva no interior de Angola. Na conferência internacional de Otto (1884 e 1885) foram elaboradas fronteiras de Angola, para garantir a proteção contra investidas de outros países europeus (sobretudo Inglaterra e França). Angola tornou-se colónia oficial portuguesa em 1886 (Stead & Rorison 2009).

Mais tarde, com a independência concedida por países europeus a outras colónias Africanas, começaram a surgir movimentos em Angola com o objetivo de tornar o país independente. Em 1953 surgiu o primeiro movimento em Angola em busca da independência, denominado por Partido da Luta Unida dos Africanos de Angola. Dois anos depois, em 1955, foi criado o Partido Comunista Angolano, seguido do Movimento Popular de Libertação de Angola, em 1956. Apenas em 1974, com a ascensão do novo regime político em Portugal governado pelo General António de Spínola, foi iniciado o processo de independência das colónias portuguesas em África. Angola tornou-se independente a Novembro de 1975 (Stead & Rorison 2009).

c. Angola de hoje

Atualmente, para além das minorias existentes dos povos Khoe-san, são identificados nove principais grupos de origem Bantu em Angola: os Ovambo, os Herero, os Xindonga, os Ganguela, os Ovimbundu, os Nhaneca-Humbe, os Bakongo, os Ambundo e os Chokwe (Figura 4).

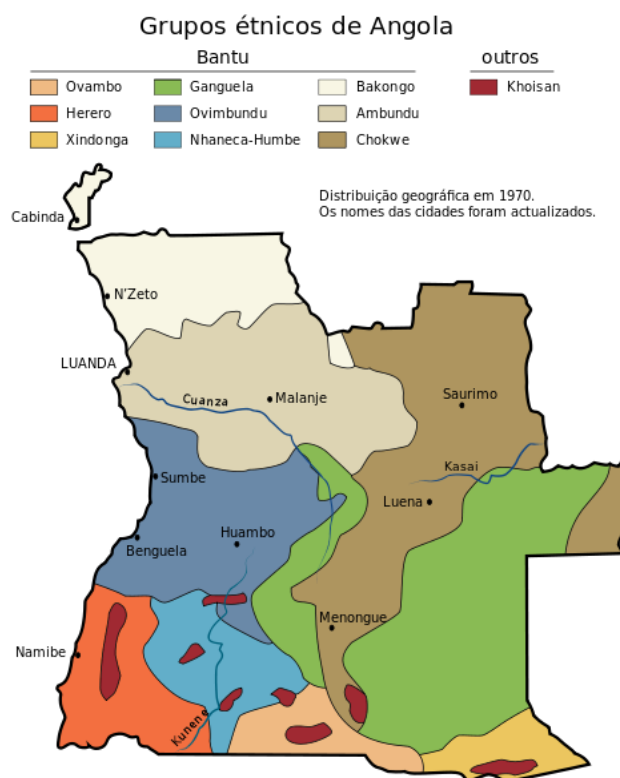


Figura 4 – Distribuição das tribos Bantu no território Angolano no ano de 1970
 (http://upload.wikimedia.org/wikipedia/commons/thumb/7/70/Angola_Ethnic_map_1970-pt.svg/488px-Angola_Ethnic_map_1970-pt.svg.png, visualizado a Outubro 2013)

Quanto ao sistema de organização política do Estado, atualmente, Angola é uma República. A população encontra-se distribuída principalmente pelas regiões urbanas de Luanda, Huambo, Lubango e Benguela. As regiões Este e Sul do país estão maioritariamente desabitadas, sendo Angola um dos países com maior área desabitada em África. Atualmente a população Angolana perfaz cerca de 16-18 milhões de indivíduos. Estima-se que aproximadamente 46% dos indivíduos tem idade inferior a 15 anos e a esperança média de vida é de 42 anos.

A língua oficial é o português, sendo utilizada pelo governo e ensinada nas escolas; contudo existem outras 42 línguas, sendo as mais comuns as línguas Bantu Umbundo, Kibundo e Kicongo. Nas regiões urbanas a maioria dos indivíduos utiliza português como língua corrente, tendo apenas algumas noções de linguagem Bantu. Por outro lado, nas regiões rurais, passa-se o contrário, onde muitas pessoas falam bem Bantu mas não dominam o português (Stead & Rorison 2009).

2.3 O fenómeno de Migração em Portugal

Desde o início do século XX que Portugal é considerado um país de emigrantes. Este panorama alterou-se por volta dos anos 70, com o crescente fluxo de imigrantes, amplamente relacionado com a fase pós-colonial dos países Africanos de língua portuguesa (Angola, Cabo Verde, Guiné-Bissau, Moçambique e São Tomé e Príncipe). Portugal foi o país escolhido por um elevado número de Africanos que fugiam às guerras civis dos seus países de origem e/ou que procuravam melhores condições de vida. Estes indivíduos imigrantes vieram contribuir para um aumento da heterogeneidade em termos sociais, culturais, religiosos, linguísticos e antropológicos em Portugal (Mendes *et al.* 2011).

Segundo dados estatísticos do Serviço de Estrangeiros e Fronteiras - SEF - referentes ao ano de 2013, o número total de imigrantes a residir em Portugal atingiu os 401 320 indivíduos. As nacionalidades mais representativas deste valor eram, por ordem decrescente, Brasil (92 120), Cabo Verde (42 401), Ucrânia (41 091), Roménia (34 204), Angola (20 177), China (18 637), Guiné-Bissau (17 846), Reino Unido (16 471), São Tomé e Príncipe (10 304) e Moldávia (9 971) (Serviço de Estrangeiros e Fronteiras 2013) (Figura 5). De forma geral, verificava-se uma distribuição territorial predominante na região litoral do país, especialmente em Lisboa, Faro e Setúbal.

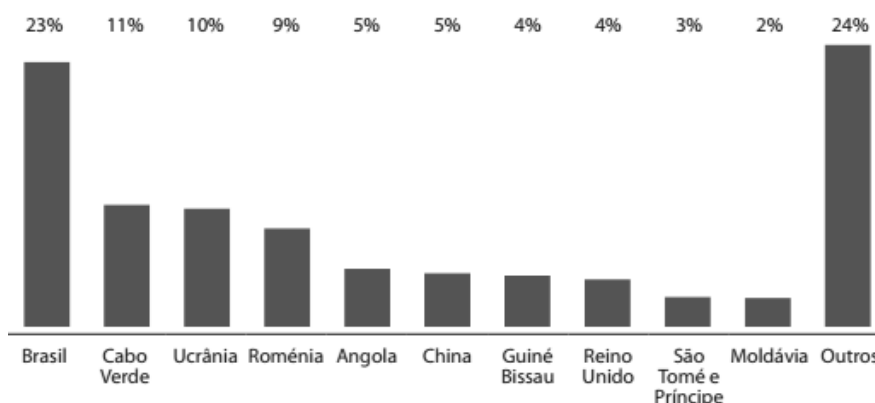


Figura 5 - Percentagem da população imigrante em Portugal no ano de 2012 (Serviço de Estrangeiros e Fronteiras 2013)

Nos anos 80 e 90, verificou-se um aumento do fluxo migratório de imigrantes Angolanos em Portugal, que tinham como principal motivação a procura de trabalho. De acordo com dados estatísticos do ano de 2013, verifica-se uma concentração dos indivíduos imigrantes desta nacionalidade pelos distritos de Lisboa, Setúbal e Porto. Somente na região de grande Lisboa, no final de 2013, residiam 12561 indivíduos de nacionalidade Angolana (Serviço de Estrangeiros e Fronteiras 2013; Anon 2014). Dados relativos ao ano de 2009 indicavam uma distribuição de indivíduos Angolanos por sexo equilibrada, com um número ligeiramente superior de mulheres Angolanas a residir em Portugal. Relativamente à idade dos indivíduos, cerca de 91% tinham idade inferior a 45 anos (Mendes *et al.* 2011).

2.4 Mitocôndria

2.4.1 Estrutura da mitocôndria

A mitocôndria é um organelo celular presente no citoplasma das células eucariotas. Estes organelos podem variar em forma e tamanho, contudo, geralmente apresentam estrutura cilíndrica e alongada, com diâmetro entre 0,5µm e 1µm. São compostos por uma membrana externa e uma membrana interna. A membrana externa apresenta uma proporção igual de proteínas e lípidos (50:50); possui ainda canais transmembranares compostos por porinas, que permitem a passagem de íons e moléculas até 10 Kdaltons de peso molecular. A membrana interna, por sua vez, apresenta uma constituição mais rica em proteínas que lípidos (80:20) e é mais impermeável a pequenas moléculas e íons; porém fazem também parte da sua constituição proteínas transportadoras que contribuem para a passagem de metabolitos como ATP (Trifosfato de Adenosina) e ADP (Difosfato de Adenosina) (Voet *et al.* 2012; Lodish *et al.* 2000). Esta última membrana apresenta várias invaginações, dobrando-se numa série de sulcos internos denominados por crista que delimitam dois compartimentos mitocondriais: a matriz (circunscrita pela membrana interna) e o espaço intermembranar (entre a membrana interna e externa) (Alberts *et al.* 2010).

2.4.2 Função da mitocôndria

A produção de ATP através do processo de fosforilação oxidativa é a principal função da mitocôndria. Moléculas de ATP são geradas a partir de NADH (*Nicotinamide adenine dinucleotide*) ou de FADH₂ (*Flavin adenine dinucleotide*). O NADH e o FADH₂ são moléculas ricas em energia uma vez que contêm um par de elétrons com elevado potencial de transferência. Quando estes elétrons são utilizados para reduzir oxigénio a água através de vários passos de transferência eletrônica, uma elevada quantidade de energia é libertada, utilizada para produzir ATP. Neste processo estão envolvidos quatro complexos enzimáticos – complexos I, II, III e IV - localizados na membrana interna da mitocôndria sendo todos eles transmembranares, à exceção do complexo II. O fluxo de elétrons produzidos pelo NADH e FADH₂ através destes complexos induz um bombeamento de prótons da matriz para o espaço intermembranar. A molécula de ATP é sintetizada quando um próton flui de volta para a matriz mitocondrial através do complexo ATP sintase (Lodish *et al.* 2000; Nelson & Cox 2004).

2.4.3 Origem da mitocôndria e Hipótese endossimbiótica

A teoria que explica o fenómeno de incorporação das mitocôndrias no interior das células eucariotas tem por base um processo de endossimbiose.

Os primeiros organismos com metabolismo aeróbio terão sido as bactérias. A teoria endossimbiótica postula que pequenas bactérias aeróbias foram integradas e sobreviveram no interior de células eucariotas primitivas. Estas, que viviam de forma anaeróbia, terão adquirido

a capacidade de realizar a fosforilação oxidativa ao estabelecer uma relação de simbiose com a bactéria a viver no seu citoplasma (Nelson & Cox 2004). A relação de simbiose terá garantido às bactérias a sua sobrevivência apesar da exposição a condições adversas, como o aumento da concentração de elementos altamente reativos e potencialmente tóxicos no meio ambiente. Ao longo do tempo, o organismo endossimbionte ter-se-á tornado parte integral da célula eucariota primitiva, desenvolvendo-se e dando origem à mitocôndria. O núcleo das células dos atuais seres eucariotas terá derivado de genes presentes no organismo eucariota primitivo. A maior parte dos genes da mitocôndria primitiva terão sido transferidos para o núcleo celular; no entanto, um pequeno número de genes mitocondriais terá permanecido no organelo, fazendo parte do genoma mitocondrial atual (Nelson & Cox 2004; Lodish *et al.* 2000). A maioria dos polipéptidos mitocondriais são codificados no genoma nuclear, sintetizados no citoplasma da célula e importados para a mitocôndria após transcrição. As proteínas translocases, presentes na membrana externa, auxiliam o transporte dos polipéptidos até à mitocôndria (Lodish *et al.* 2000). Apesar da transferência de grande parte dos genes mitocondriais para o núcleo celular, a mitocôndria manteve a capacidade de se multiplicar ou fundir de forma independente da célula hospedeira (Rapley & Whitehouse 2007; Cummins 1998).

São diversas as evidências que suportam a teoria endossimbiótica como explicação da incorporação das mitocôndrias nos seres eucariotas. Muitos seres eucariotas unicelulares são hospedeiros de bactérias endossimbióticas. Por outro lado, as mitocôndrias apresentam tamanho semelhante às bactérias e replicam o seu próprio ADN, que apresenta semelhanças com o ADN das bactérias. A mitocôndria tem a capacidade de processar ribossomas, alguns deles de tamanho e estrutura semelhante aos ribossomas das bactérias. Por fim, os antibióticos que inibem a síntese proteica nas bactérias, mas não afetam a síntese proteica em eucariotas, produzem o mesmo efeito na mitocôndria (Pierce 2012).

As mitocôndrias são organelos semiautónomos que mantem uma relação de dependência com o hospedeiro. Células com mitocôndrias dependem destes organelos para realizarem a fosforilação oxidativa e, estes, por sua vez dependem da célula para garantir a sua própria existência.

2.5 ADN mitocondrial

O genoma humano corresponde à totalidade da constituição genética do indivíduo e é formado por uma componente nuclear e uma componente mitocondrial.

O ADN nuclear (ADNn), presente no interior do núcleo das células, está organizado em cromossomas, integrando proteínas denominadas por histonas. Cada célula, por sua vez, é constituída por 22 pares de cromossomas autossómicos e dois cromossomas sexuais (cromossoma X e cromossoma Y no homem e dois cromossomas X na mulher) (Makaowski 2001). Assim, cada célula possui um total de 46 cromossomas (Regateiro 2007; Pinheiro 2010).

O ADN mitocondrial (ADNmt) é constituído por uma molécula circular localizada no interior das mitocôndrias. O genoma mitocondrial representa apenas uma ínfima fração da totalidade do genoma de um indivíduo.

Estes dois elementos do genoma humano, nuclear e mitocondrial, apresentam particularidades distintas, apresentadas na Tabela 1.

Tabela 1 – Comparação entre o ADN nuclear e o ADN mitocondrial (adaptado de Butler 2011).

Características	ADN mitocondrial	ADN nuclear
Localização	Mitocôndria	Núcleo
Estrutura	Cadeia dupla circular	Dupla hélice
Nº de cópias por células	Variado, podendo ser >1000	2
Funcionamento	Semi-autónomo	Autónomo
Herança	Materna	Materna e paterna
Característica do genoma	Haplóide	Diplóide
Recombinação	Não	Sim
Mecanismo de reparação	Não	Sim
Taxa de mutação	Elevada (5-10 vezes superior à do nuclear)	Baixa
Poder de individualização	Partilhado pela linhagem materna	Único para cada indivíduo (exeto gémeos homozigóticos)
Sequência de Referência	Descrita em 1981 (Anderson e colaboradores)	Descrita em 2001 (Projeto do Genoma Humano)

2.5.1 Estrutura do ADNmt

O ADN mitocondrial, descoberto por Margit M. K. Nass e Sylvan Nass nos anos 60 através de microscopia eletrónica (Nass & Nass 1963), é uma molécula polinucleotídica organizada em cadeia dupla circular e não sujeita ao processo de *crossing over*.

Uma molécula de ADNmt é composta por aproximadamente 16 569 pares de base (pb), contudo este número pode variar devido a inserções e/ou deleções que possam ocorrer na sua sequência. As duas cadeias polinucleotídicas que compõem a molécula circular de

ADNmt, são diferenciadas uma da outra pela composição nucleotídica de guaninas. A cadeia polinucleotídica com uma contribuição superior das purinas Guaninas (G), é denominada por cadeia pesada, em inglês *heavy* (H), por apresentar um maior peso molecular. A cadeia leve, ou *light* (L), apresenta maior percentagem de pirimidinas e, por isso, um menor peso molecular.

A molécula de ADNmt é composta por duas regiões, identificadas de acordo com a função que desempenham: a região codificante e a região não codificante.

A região codificante representa 90% do genoma mitocondrial. Inclui 37 genes, dos quais 13 codificam subunidades de proteínas da cadeia respiratória utilizadas no processo de fosforilação oxidativa e produção de energia, 22 genes codificam para ARN de transferência e 2 genes codificam para ARN ribossómico (Butler 2005; Pinheiro 2010). No entanto, a maioria das proteínas mitocondriais (cerca de 1 100 diferentes tipos) é codificada por genes nucleares, sintetizada em ribossomas citoplasmáticos e importada até à mitocôndria (Nelson & Cox 2004).

A região não codificante não codifica para nenhum produto de gene. É também denominada por região “laço de deslocamento” (do inglês *displacement loop* ou *D-loop*), definição associada à presença da origem de replicação da cadeia pesada, ou por região controlo, por possuir os promotores de transcrição de ambas as cadeias polinucleotídicas. A região não codificante tem uma extensão de 1 122 pb e é nela que estão presentes os segmentos de maior variabilidade genética do genoma mitocondrial denominados por segmentos hipervariáveis – HVI, HVII e HVIII (*HiperVariable Region* I, II e III) (Figura 6).

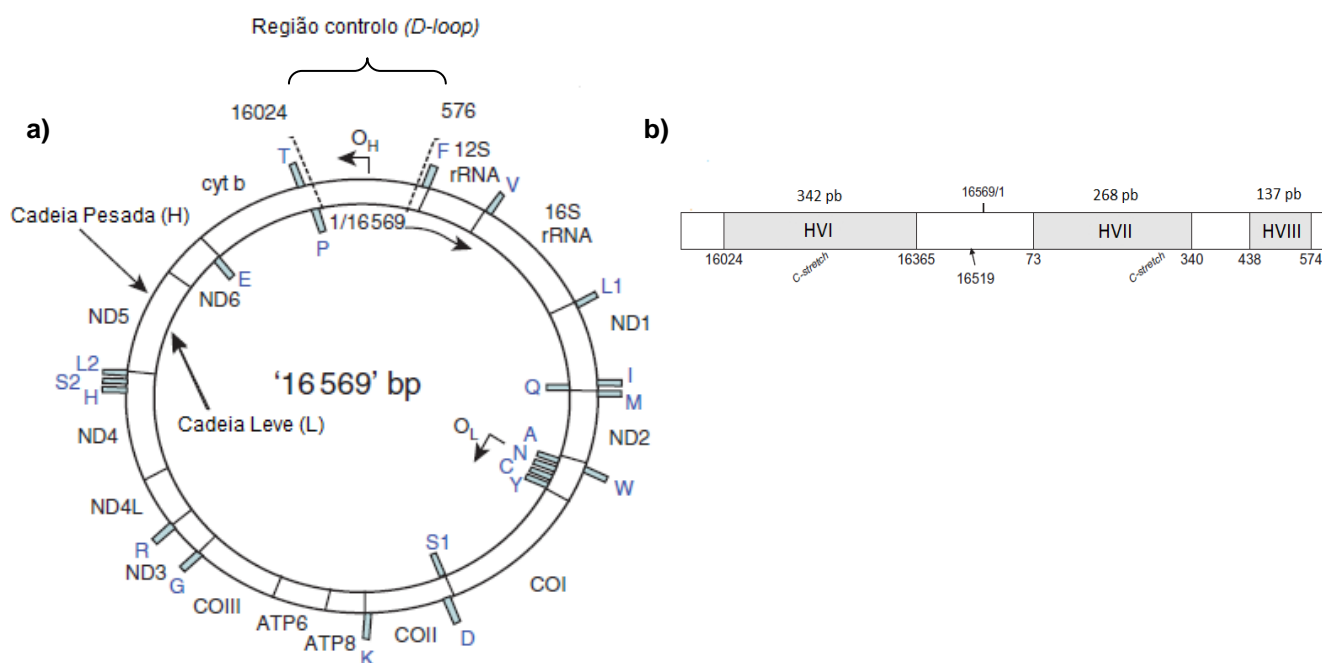


Figura 6 – Estrutura do ADNmt. a) Molécula de ADNmt; b) Região não codificante do ADNmt (adaptado de Butler 2011)

Apesar de existir interação entre os genomas mitocondrial e nuclear, a replicação do ADNmt é independente do ciclo de divisão nuclear. O ADNmt replica através do processo de replicação assimétrica onde as duas cadeias de ADNmt não se replicam de forma sincronizada. Existem duas origens distintas para replicação do ADNmt: a origem de replicação da cadeia pesada (O_H), localizada na região controlo da molécula, e a origem de replicação da cadeia leve (O_L), localizada entre as posições 5 721 e 5 798 do ADNmt.

Nas células humanas, cada cadeia de ADNmt contém uma região promotora para o início da transcrição: o promotor da cadeia leve - LSP (*light-strand promotor*) - e o promotor da cadeia pesada - HSP (*heavy-strand promotor*). A transcrição produz percursos policistronicos de ARN, que contêm toda a informação genética presente em cada uma das cadeias. Estes transcritos primários são processados de forma a dar origem aos ARNs mensageiros, aos ARNs ribossómicos e aos ARNs de transferência (Falkenberg *et al.* 2007).

O código genético mitocondrial é muito semelhante ao código nuclear, contudo apresenta algumas diferenças. Fazem parte da constituição desta molécula as purinas adenina e guanina e as pirimidinas citosina e timina. As diferenças residem nas codificações dos codões presentes nas sequências de ADNmt. Dos quatro codões *STOP* que apresenta, dois deles – UAA e UAG – também servem como codão stop no código genético nuclear; os outros dois – AGA e AGG – codificam arginina no ADNn. O codão UGA não é lido como sendo um codão STOP mas sim como o aminoácido triptofano; o codão AUA é uma metionina e não uma isoleucina; e o codão de iniciação AUG por vezes é substituído por AUA ou AUU em genes mitocondriais. Por outro lado, o genoma mitocondrial não é universal para todas as mitocôndrias, variando de organismo para organismo (Lewin 2008).

2.5.2 Características do ADNmt

Ao longo das últimas décadas o ADNmt tornou-se o marcador genético de eleição em estudos de evolução humana e populacional, assim como em estudo de casos forenses e clínicos. O interesse na análise do ADNmt surgiu pelas características da molécula: o elevado número de cópias por célula num mesmo indivíduo, a herança uniparental materna, a ausência de recombinação e a elevada taxa de mutação.

a. Elevado número de cópias

Contrariamente ao que acontece com o ADNn, com duas cópias numa célula, a molécula de ADNmt apresenta centenas, ou mesmo milhares de cópias.

Cada mitocôndria contém no seu interior, em média, 4 a 5 cópias de ADN mitocondrial. Por sua vez, uma única célula pode conter centenas de mitocôndrias no citoplasma, completando centenas, ou até mesmo milhares de cópias de ADNmt (Iborra *et al.* 2004; Butler 2005). Estas moléculas existem em maior número em tecidos que exigem uma maior quantidade de oxigénio, como por exemplo no cérebro e no músculo-esquelético.

b. Herança materna

Uma outra característica que torna o ADNmt útil e importante, tanto em estudos de casuística forense, como em estudos de linhagens, ancestralidade e evolução de populações, é a herança materna.

Indivíduos da espécie humana herdam todo o seu genoma mitocondrial através do parente feminino. Excetuando possíveis mutações, o ADNmt de indivíduos da mesma linhagem materna, será idêntico (Figura 7).

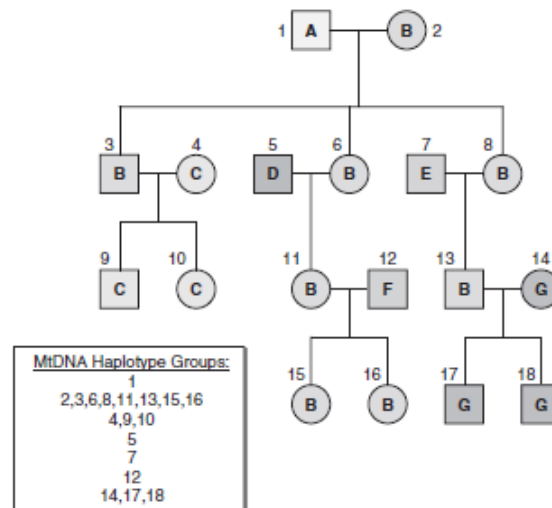


Figura 7 – Representação da transmissão materna do ADNmt referente a 18 indivíduos (Butler 2005). Os círculos representam indivíduos do sexo feminino e os quadrados indivíduos do sexo masculino.

Existem diversas explicações para a existência da transmissão uniparental. Por um lado, esta pode estar relacionada com a diferença entre o número de mitocôndrias presentes no oócito e no espermatozoide; enquanto os oócitos apresentam dimensão superior e contêm entre 10^5 a 10^6 mitocôndrias, os espermatozoides são menores e contêm entre 10^2 a 10^3 mitocôndrias. Desta forma, verifica-se um fenómeno de diluição resultante do excesso de mitocôndrias presentes no oócito comparativamente ao espermatozoide (Holland & Parsons 1999). Por outro lado, durante a concepção, considerando que apenas a cabeça do espermatozoide entra no interior oócito, haverá a subsequente fusão do núcleo do espermatozoide com o núcleo do oócito. Quando a divisão celular ocorre, com a divisão do zigoto e formação do blastocisto, todas as partes celulares, exceto o núcleo, são consistentes com o oócito materno. Ainda, diversos autores confirmam que as mitocôndrias pertencentes à linhagem paterna que possam ser introduzidas no oócito são destruídas por atividade proteolítica endógena durante a embriogénese, uma vez que o ADNmt paterno está marcado para destruição por ubiquitinação (Cummins 1998; Pakendorf & Stoneking 2005; Sutovsky *et al.* 2000; Rapley & Whitehouse 2007).

Desta forma, as mitocôndrias paternas parecem desaparecer numa fase inicial da embriogénese por seleção, por inativação ou simplesmente por diluição pelo excesso de mitocôndrias presentes do oócito.

Existem casos publicados referentes à transmissão paterna do genoma mitocondrial. A primeira evidência de transmissão paterna de ADNmt foi reportada em 1999 por Hagelberg e colaboradores (Hagelberg *et al.* 1999). No seu estudo Hagelberg observou a presença de uma mutação em indivíduos de três linhagens de ADNmt não relacionadas de uma população. A presença desta mutação rara em três haplogrupos distintos numa região geográfica isolada é extremamente improvável, o que poderia ser explicado, segundo os autores, por transferência paterna e recombinação. Mais tarde, no ano de 2002 foi reportado por Schwartz e Vissing o caso de um indivíduo que continha ADNmt de origem paterna (Schwartz & Vissing 2002). O indivíduo apresentava uma miopatia resultante da deleção de 2pb – 5132delAA – no gene ND2 do ADNmt, responsável por codificar uma subunidade de enzima do complexo I da cadeia respiratória da mitocôndria. Através do seu estudo, Schwartz e Vissing, determinaram que a mutação tinha origem paterna, uma vez que o haplótipo pertencente ao indivíduo em estudo coincidia com o haplótipo de ADNmt do tio e pai do indivíduo (à exceção da deleção referida). Desta forma, esta mutação de 2pb teria surgido espontaneamente através da linhagem paterna. Contudo, desde então, não foram reportados mais casos referentes à existência de herança paterna de ADNmt (Shoubridge 2000; Martin & Muller 2007).

c. Ausência de recombinação

Outra importante característica do ADNmt é a ausência de recombinação genética, o que associa a origem desta molécula num único antepassado comum feminino. Uma vez que as mutações que ocorrem na linhagem materna são herdadas pelos descendentes, o ADNmt contém informação importante relativa ao seu ascendente materno.

Diversos estudos tentaram demonstrar a ocorrência de recombinação em mitocôndrias animais (Awadalla *et al.* 1999; Hagelberg *et al.* 1999), contudo estes foram posteriormente criticados por não serem reprodutíveis.

Na presença de recombinação genética não seria possível estabelecer relações filogenéticas de ADNmt. O fenómeno de ausência de recombinação explica que a variabilidade existente no genoma mitocondrial entre diferentes indivíduos é originada pela existência de mutações, que se vão acumulando ao longo do tempo.

d. Elevada taxa de mutação

Uma mutação é definida como uma alteração permanente na sequência de ADN. A taxa de mutação é a frequência com que uma sequência é alterada da forma *wild type* para a forma mutada. Por sua vez, a frequência de mutação é definida como a incidência de um tipo específico de mutação num grupo de indivíduos. A taxa de mutação varia de acordo com o local do genoma em que ocorre, sendo este valor 5 a 10 vezes superior no ADNmt que no ADNn. A taxa de mutação do ADNmt varia entre 0.075×10^{-6} e 0.165×10^{-6} substituições/posição/ano (Pakendorf & Stoneking 2005) o que origina múltiplos polimorfismos de ADNmt de frequência variada em diversas populações (Regateiro 2007). A taxa de mutação pode ser influenciada por três principais fatores: a frequência com que esta alteração primária

acontece no ADN; a probabilidade dessa alteração ser reparada; e a probabilidade da mutação ser reconhecida e mantida.

Na região controlo do ADNmt a taxa de mutação é relativamente heterogénea, apresentando *hotspots* mutacionais, que mutam quatro a cinco vezes mais rapidamente que qualquer outro local dessa mesma região. Nestes *hotspots* mutacionais a probabilidade de ocorrência de mutações é, portanto, mais elevada (Rapley & Whitehouse 2007; Pakendorf & Stoneking 2005).

As mutações podem ser originadas por fatores intrínsecos ou fatores extrínsecos. Os primeiros, que resultam de alterações naturais na estrutura do ADNmt, dão origem a mutações espontâneas. A extensão da cadeia de ADNmt e o tipo de base presente em cada posição nucleotídica são características que podem afetar a incidência de mutações espontâneas. Por outro lado, as mutações que resultam de agentes extrínsecos (químicos ambientais ou radiação) são denominadas por mutações induzidas. As mutações induzidas são, geralmente, mais comuns que as mutações espontâneas.

A cadeia respiratória é o maior produtor de espécies reativas de oxigénio nas células. A proximidade com a cadeia de transferência eletrónica e a ausência de histonas protetoras, tornam o genoma mitocondrial vulnerável ao ataque oxidativo por espécies reativas de oxigénio (Cummins 1998; Nelson & Cox 2004). Por outro lado, a baixa fidelidade da ADNmt polimerase torna o sistema de replicação de ADNmt menos eficiente a corrigir erros e a reparar mutações, comparativamente ao nuclear (Rapley & Whitehouse 2007; Nelson & Cox 2004; Johnson & Johnson 2001).

As mutações podem ser substituições, deleções ou inserções. Devido à complementaridade entre bases azotadas das duas cadeias de ADNmt, quando uma base de uma das cadeias é alterada, a base da cadeia correspondente irá também sofrer alteração, na próxima fase de replicação.

As substituições podem ser de dois tipos, substituição por transição ou substituição por transversão. No primeiro caso, uma purina é substituída por outra purina ou uma pirimidina é substituída por outra pirimidina. Na substituição por transversão, uma pirimidina é substituída por uma purina ou vice-versa (Malyarchuk *et al.* 2002).

As deleções e inserções são tipos de mutações que correspondem à remoção ou adição, respetivamente, de um ou mais pares de bases (Pierce 2012).

O fenómeno de *bottleneck* ou afunilamento genético define a taxa de fixação das mutações nos indivíduos. Uma população que atravessa uma fase de *bottleneck*, expandindo-se a partir de um genótipo fundador predominante, apresentará uma variabilidade genética reduzida (Howell & Smejkal 2000). Este fenómeno irá resultar numa rápida e aleatória deriva genética, com perda de algumas mutações e a uma rápida fixação de outras. Se as mutações

fixadas forem fracamente patogénicas, vantajosas ou neutras, irão contribuir para a evolução molecular do genoma mitocondrial em descendentes maternos (Shoubridge 2000; Howell & Smejkal 2000).

2.5.3 Sequência de Referência de Cambridge - CRS

A primeira sequência de ADNmt estudada na totalidade foi obtida no laboratório de Frederick Sanger em Cambridge e publicada em 1981 por Anderson e colaboradores (Anderson *et al.* 1981), tendo sido designada por Sequência de Referência de Cambridge (*Cambridge Reference Sequence* – CRS). No seu estudo, Anderson e colaboradores (1981), atribuíram uma numeração aos nucleótidos de acordo com a sua posição na sequência; esta classificação inicia-se na posição 1 junto à origem de replicação da cadeia pesada e termina na posição 16 569. A região controlo do ADNmt encontra-se entre as posições 16 024 e 576. Relativamente aos segmentos hipervariáveis desta região, HVI estende-se da posição 16 024 à posição 16 365, HVII da posição 73 à posição 340 e HVIII da posição 430 à posição 576.

Estudos posteriores verificaram a existência de incongruências na sequência publicada por Anderson e colaboradores (1981). Consequentemente, em 1999, Andrews e colaboradores, publicaram a mesma sequência de ADNmt revista, onde foram identificados e corrigidos 11 erros. Desta forma a sequência passou a ser denominada como Sequência de Referência de Cambridge revista (*Revised Cambridge Reference Sequence* – rCRS) (Andrews *et al.* 1999).

O conjunto das alterações presente numa sequência de ADNmt, relativamente à rCRS, é denominado por haplótipo. Assim, os polimorfismos de ADNmt são identificados e classificados pela posição em que ocorrem seguida do nucleótido que deu origem a essa mesma alteração. Na construção do haplótipo está implícito que as posições nucleotídicas não identificadas serão idênticas à rCRS (Pinheiro 2010). O conjunto dos polimorfismos específicos de ADNmt forma grupos de linhagens próximas umas das outras, dando origem aos haplogrupos e subhaplogrupos, classificados com letras do alfabeto. Sequências pertencentes a um determinado haplogrupo ou subhaplogrupo apresentam alterações específicas, representativas de um ancestral comum (Pinheiro 2010).

A amostra de ADN estudada pelos grupos de investigação de Anderson e de Andrews pertencia a um indivíduo de origem materna Europeia, com sequência classificada no haplogrupo H. Apesar da sequência não ser representativa da maioria do ADNmt presente nos indivíduos (notável na posição 263 e 315), continua a ser utilizada para o estudo da região controlo do ADNmt. A sequência encontra-se alojada no GenBak com a identificação J01415.2 e os resultados publicados são referentes à cadeia leve do ADNmt estudado.

2.5.4 Eva mitocondrial

A transmissão materna e a ausência de recombinação inerentes ao ADNmt permitem estimar a data do antepassado comum mais recente do Homem moderno, em inglês *Most Recent Common Ancestor* (MRCA), que transmitiu a sua sequência de ADNmt a todos os indivíduos atuais. A partir do estudo do ADNmt é possível obter informação relativa ao antepassado comum feminino a todos os indivíduos, denominado por Eva mitocondrial, em reconhecimento da linhagem materna do ADNmt. A sequência de ADNmt deste antepassado terá sido transmitida para a população humana atual. Corresponde ao genoma mitocondrial de um indivíduo do sexo feminino, exposto a condições reprodutivas favoráveis, levando a um aumento de frequência do seu ADNmt comparativamente aos restantes. A Eva mitocondrial expressa a ideia de que todos os humanos modernos derivam de um antepassado comum que viveu em África entre 100 000 a 200 000 anos atrás. Esta teoria suporta o pressuposto relativo à origem do Homem em África (Lewin 2008).

O estudo científico que determinou o tempo e lugar do antepassado comum mais recente foi desenvolvido no ano de 1987, por Allan Wilson e seus colaboradores que estudaram os haplótipos de ADNmt de 147 indivíduos, com o objetivo de construir uma árvore filogenética que extrapolava até um antepassado comum (Cann *et al.* 1987).

2.5.5 Heteroplasmia

Na análise do ADNmt existem diversos fatores que devem ser considerados, de forma a garantir a correta obtenção e interpretação de resultados. Um desses fatores é a possibilidade da presença de mais do que uma sequência de ADNmt no mesmo indivíduo, fenómeno que pode dar origem a resultados ambíguos. A condição caracterizada pela coexistência de duas ou mais populações de ADNmt, com diferente composição nucleotídica, numa mesma mitocôndria, célula, tecido ou indivíduo, é denominada por heteroplasmia (Brandstätter & Parson 2003). Na homoplasmia todas as cópias do genoma mitocondrial são idênticas (Taylor & Turnbull 2007). Desta forma é possível distinguir as mutações que afetam todas as cópias do genoma mitocondrial – mutações homoplasmicas – das que apenas estão presentes em algumas cópias de ADNmt – mutações heteroplasmicas (Taylor & Turnbull 2007; Comas *et al.* 1995).

Os primeiros estudos realizados com base na análise de heteroplasmias do ADNmt estavam associados a patologias mitocondriais. Contudo, investigações posteriores revelaram que este fenómeno está presente na população em geral (Calloway *et al.* 2000; Butler 2005; de Camargo *et al.* 2011). Este reconhecimento e deteção de diferentes níveis de heteroplasmia em indivíduos saudáveis pode ser explicado pelo aumento da sensibilidade das técnicas utilizadas nos dias de hoje, ou mesmo pelo tipo de tecido utilizados nos diferentes estudos.

O “destino” das mutações heteroplasmicas que atingem o ADNmt, isto é, a sua continuidade, ou não, no genoma mitocondrial, depende de diversos factores, entre os quais o tipo e localização desses polimorfismos e a taxa de replicação da célula.

Existem dois tipos de heteroplasma distintos – heteroplasma de sequência ou posição e heteroplasma de comprimento. A primeira está associada a substituições nucleotídicas numa determinada posição, presente em apenas algumas sequências de ADNmt de um indivíduo e é caracterizada pela sobreposição de dois picos no eletroferograma obtido por sequenciação. A heteroplasma de comprimento resulta de mutações que envolvem inserções ou deleções de nucleótidos na sequência de ADNmt e é o tipo de heteroplasma mais comum. (Butler 2005; Taylor & Turnbull 2007). A heteroplasma de comprimento acontece, preferencialmente, nas zonas polimórficas compostas por Citosinas (C) dos segmentos HVI, HVII e HVIII da região controlo do ADNmt. Estas regiões de repetições mononucleotídicas são conhecidas como *hotspots* mutacionais que dão origem ao fenómeno de heteroplasma de comprimento.

Em 2000, Calloway e colaboradores determinaram que poderia existir uma relação entre a frequência da ocorrência de heteroplasma e a idade, verificando-se um aumento da frequência de ocorrência de heteroplasma com o aumento da idade (Calloway et al. 2000). Contudo, estudos posteriores (de Camargo *et al.* 2011; Sondheim *et al.* 2011) vieram contrariar esta ideia, demonstrando que a frequência de ocorrência de heteroplasma se mantém constante ao longo da vida, havendo uma distribuição igual entre os diferentes grupos etários. De igual forma, tanto o sexo dos indivíduos como o haplogrupo em que o haplótipo de ADNmt se inclui, não fornecem informações significativas relativamente à frequência de ocorrência de heteroplasma. Contrariamente, o fenómeno da heteroplasma não é uniforme em todos os tecidos e células (Sondheim *et al.* 2011; Samuels *et al.* 2013), estando demonstrado que é mais comum no músculo-esquelético e no cérebro (Jazin *et al.* 1996). A frequência de ocorrência de heteroplasma também varia entre diferentes cabelos do mesmo indivíduo (Taylor & Turnbull 2007; Lutz *et al.* 1996; Butler 2005; Roberts & Calloway 2011; Alonso *et al.* 2002).

A presença de duas ou mais sequências de ADNmt num mesmo indivíduo pode acontecer a três níveis. Por um lado, ao nível da célula, sendo que uma única célula contém mitocôndrias homoplasmicas, mas diferentes células contêm diferentes variantes; por outro lado ao nível da mitocôndria, em que uma célula pode conter diferentes haplótipos de ADNmt, mas cada mitocôndria contém o seu homoplasmico; por fim, ao nível dos ácidos nucleicos, onde uma mitocôndria contém diferentes tipos de ADNmt.

Quando ocorre a divisão celular de uma célula heteroplásmica, o ADNmt mutado e não mutado pode ser distribuído de forma aleatória para as células filhas. Este fenómeno é denominado por segregação replicativa. Como resultado a percentagem de moléculas mutadas e de moléculas *wild type* pode variar entre células. Quando este fenómeno de segregação replicativa ocorre em células somáticas, pode criar variação fenotípica num único organismo fazendo com que o fenótipo seja influenciado pela proporção de ADNmt que apresenta a mutação nos diferentes tecidos (Taylor & Turnbull 2007).

2.5.6 Aplicações do genoma mitocondrial

O estudo do genoma mitocondrial tem aplicação em três principais áreas: a genética médica, através da análise de mutações que se traduzem em doenças, uma vez que um elevado número de doenças hereditárias são causadas por mutações em genes da região codificante da molécula de ADNmt; a genética populacional (ou antropologia molecular), através do estudo das diferenças e das similaridades entre as populações humanas; e a genética forense, pela obtenção de haplótipos de ADNmt e verificação da coincidência entre duas ou mais amostras biológicas.

Os primeiros estudos realizados na molécula de ADNmt humano tinham por base o uso de técnicas de restrição enzimática - *Restriction fragment length polymorphism* (RFLPs) – com a utilização de enzimas de restrição na totalidade do genoma mitocondrial (Brown 2006). O avanço tecnológico de métodos de amplificação - *Polymerase Chain Reaction* (PCR) - e sequenciação - sequenciação de Sanger (Sanger & Nicklen 1977) - de ADN permitiram o estudo mais detalhado da molécula de ADNmt. Neste sentido, começou-se por sequenciar o segmento HVI da região controlo de ADNmt, entre as posições 16 024 e 16 365 (Castrì et al. 2009; Nagai et al. 2003). Mais tarde, estudos incluíram a sequenciação dos segmentos HVI e HVII da região controlo do ADNmt, entre as posições 16 024 e 340 (Afonso Costa et al. 2010; Alvarez et al. 2007; Lima et al. 2006; Coelho et al. 2009; Pereira et al. 2000; Plaza et al. 2004; Gonçalves et al. 2003; Beleza et al. 2005). Nos últimos anos, com o desenvolvimento de tecnologias de sequenciação de alto rendimento, tornou-se comum o estudo da região controlo total do ADNmt (incluindo os segmentos HVI, HVII e HVIII) (Castro de Guerra et al. 2012; Scheible et al. 2011; Mairal et al. 2013) e da totalidade do genoma mitocondrial (Gonder et al. 2006; Knapp et al. 2012).

a. ADNmt e genética populacional

Os estudos de genética populacional permitem determinar a composição genética de uma população. Esta área debruça-se nas diferenças e similaridades genéticas que podem ser detetadas dentro de uma população ou entre populações (Hartl & Clark 1997; Relethford 2012).

O genoma mitocondrial humano tem sido utilizado em estudos de evolução humana para estimar a origem geográfica e a dinâmica das populações. É possível, através destes estudos, obter informação sobre eventos históricos da humanidade.

Os estudos da transmissão materna do ADNmt permitem complementar informações referentes à origem do Homem moderno em África e a sua dispersão pelo mundo (Pinheiro 2010; Rosa & Brehem 2011; Howell et al. 1996).

O primeiro estudo de genética populacional com recurso ao ADNmt foi realizado nos anos 80 através de técnicas de RFLPs (Brown 1980).

Os haplótipos de ADNmt inseridos num determinado haplogrupo apresentam polimorfismos específicos como consequência da partilha de um ancestral comum. Neste

sentido os haplogrupos de ADNmt podem estar associados a determinadas regiões geográficas e, em alguns casos, a grupos étnicos particulares.

As sequências de ADNmt de cada indivíduo são utilizadas para construir árvores filogenéticas (Howell *et al.* 1996). Em 2007, foi desenvolvida na base de dados MITOMAP, a primeira árvore filogenética a partir de sequências de ADNmt. Contudo esta apenas incluía as mutações presentes na região controlo do ADNmt (Ruiz-Pesini *et al.* 2007). No ano de 2009, Van Oven e Kayser desenvolveram uma árvore filogenética – *Phylotree* - que englobava os polimorfismos presentes na região controlo e na região codificante do ADNmt. A *Phylotree* encontra-se em constante atualização e foi revista pela última vez em Fevereiro de 2014 (van Oven & Kayser 2009). Os ramos mais profundos e antigos dessa árvore incluem haplogrupos característicos do continente Africano, mais especificamente da região Subsariana (Rosa & Brehem 2011). As linhagens mas antigas pertencem ao macrohaplogrupo L, que incluiu os haplogrupos L0, L1, L2, L3, L4, L5 e L6, todos eles específicos de populações Africanas. O haplogrupo L3 ramifica-se nos haplogrupos M e N, que se dispersaram para a região da Eurásia há cerca de 60 000-70 000 anos e deram origem às linhagens presentes nestas regiões. As populações Europeias derivam do haplogrupo N, onde se incluem as linhagens as linhagens H, I, J, K, T, U, V, W e X que, no conjunto, representam 98% da população Europeia. As linhagens Asiáticas derivam tanto do haplogrupo M como do haplogrupo N. No primeiro incluem-se as ramificações C, D, E e Z e no segundo as ramificações A, B, F e J. Nas populações nativas Americanas verifica-se a presença dos haplogrupos A, B, C, e D, característicos da região Asiática, o que indica a sua presença na região Americana, anterior à colonização Europeia (Pinheiro 2010) (Figura 8).

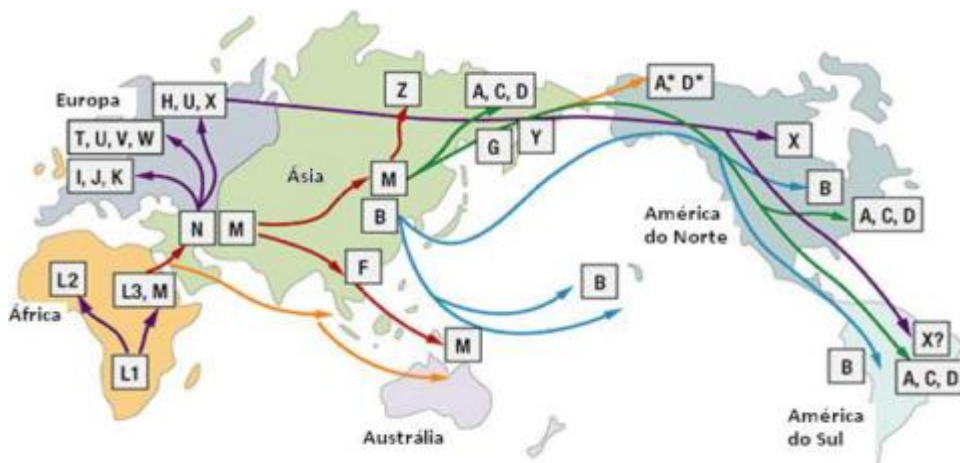


Figura 8 – Mapa representativo da dispersão do Homem por todo o mundo e distribuição geográfica dos haplogrupos de ADNmt (Adaptado de <http://www.familytreedna.com>)

África apresenta uma grande variabilidade genética que é sugestiva de uma acumulação de mutações durante um longo período de tempo. Os haplótipos europeus apresentam, em média, apenas metade das mutações características das populações Africanas (Micklos & Freyer 2003).

b. Aplicação forense do ADNmt

Nas últimas décadas a utilização de métodos analíticos que permitem o estudo de ADN, ajudaram a resolver casos em diversas investigações forenses. O estudo de STRs – *Short Tandem Repeats* – que têm por base polimorfismos presentes no ADNn, demonstra ser eficaz quando as amostras biológicas apresentam ADNn em quantidade suficiente e em qualidade desejável. Contudo, na presença de ADN antigo ou de amostras degradadas o estudo de STR revela-se inadequado. A análise do ADNmt tem-se revelado eficiente nestes casos particulares, apesar de apresentar um baixo poder de discriminação comparativamente ao ADNn (Butler 2005; Lutz *et al.* 1996; Wilson *et al.* 1995; Parson *et al.* 1998).

Em 1996 a análise de ADNmt foi realizada pela primeira vez em estudos de genética forense, num caso de violação e homicídio em Tenesse, Estados Unidos da América. Sem os resultados obtidos a partir da análise do ADNmt, apenas eram apresentadas evidências circunstanciais contra o suspeito Paul War. Após comparação dos resultados de ADNmt de diversos vestígios com resultados de ADNmt de amostra de referência do suspeito, foi reforçada a hipótese do seu envolvimento no crime (Davis 1998).

Um dos casos mais populares envolvendo o estudo do ADNmt, foi o da identificação de restos cadavéricos pertencentes a indivíduos da família Romanov. Os ossos de czar Nicholas II encontravam-se enterrados há mais de 70 anos, contudo foi obtida uma coincidência entre as sequências da região controlo do ADNmt de czar Nicholas II e de familiares da sua linhagem materna. Esta similaridade de sequências verificou-se, excetuando uma alteração heteroplasmica na posição 16169 da região HVI. Posteriormente verificou-se que essa heteroplasmia estava também presente numa amostra de ADNmt pertencente ao irmão de czar, Georgij Romanov.

A análise de ADNmt em investigações forenses, apresenta vantagens e desvantagens. Em casos onde o objetivo seja a comparação entre familiares da mesma linhagem materna, este marcador genético é adequado. Contudo, por outro lado, não é possível definir a identidade individual de cada sujeito uma vez que indivíduos da mesma linhagem materna apresentam o mesmo haplótipo (Pinheiro 2010).

Na comparação entre haplótipos de ADNmt em perícias forenses são consideradas três possíveis conclusões. A exclusão é determinada quando as sequências diferem em duas ou mais posições nucleotídicas (não incluindo heteroplasmia de comprimento) e duas amostras podem ser excluídas como tendo a mesma origem. Uma comparação inconclusiva resulta da diferença de apenas uma única posição nucleotídica entre sequências. Por outro lado, duas amostras não se podem excluir como tendo a mesma origem ou a mesma linhagem materna, se apresentarem a mesma sequência (partilham bases de ADN comum em cada posição nucleotídica, incluindo variações de comprimento comum) (Scientific Working Group on DNA Analysis Methods 2013).

Na obtenção de perfis de ADNmt de uma mostra de referência e de uma amostra de vestígio, que não se podem excluir como tendo a mesma origem, os haplótipos devem ser

introduzidos numa base de dados de ADNmt por forma a fornecer informação estatística sobre a frequência do haplótipo na população.

2.5.7 Bases de dados de ADNmt

Na comparação de haplótipos e determinação da sua frequência na população ou no cálculo da frequência de um polimorfismo específico numa determinada população, torna-se necessário a obtenção de amostras de referência de ADNmt a nível mundial.

Diversos laboratórios forenses utilizam a sua própria base de dados de ADNmt. Contudo, muitos deles não têm capacidade para gerar sequências mitocondriais em elevado número, com a obtenção de dados insuficiente. Nestes casos são utilizadas as bases de dados que se encontram descritas e referenciadas em publicações científicas da especialidade.

A base de dados SWGDAM – *Scientific Working Group for DNA Analysis Methods* – desenvolvida pelo FBI, inclui sequências de ADNmt pertencentes a indivíduos de diversas origens e grupos populacionais.

A MITOMAP (www.mitomap.org) é uma base de dados construída em 2007, que compreende os diversos polimorfismos e mutações da região controlo do ADNmt humano e inclui informação publicada e não publicada dos haplótipos de ADNmt (Ruiz-Pesini *et al.* 2007).

A base de dados mais utilizada a nível mundial é a EMPOP – *EDNAP mitochondrial DNA population database* - (www.empop.org) (Parson & Dür 2007) que é gerida e mantida pelo Instituto de Medicina Legal de Innsbrunn. Esta base de dados tem como objetivo o alojamento de sequências e o controlo da sua qualidade. Com a disponibilização de haplótipos correspondentes à região controlo total de indivíduos de todo o Mundo é possível, através da base de dados EMPOP, obter informação estatística e realizar cálculos de parâmetros populacionais importantes em estudos populacionais e forenses. A EMPOP disponibiliza os haplótipos em duas categorias. Os dados forenses são aqueles que são acompanhados por um controlo de qualidade de todas as sequências, o que permite confirmar posições nucleotídicas a qualquer momento, de qualquer amostra. Por outro lado, os dados da literatura incluem populações descritas na literatura ou ainda por publicar; neste caso, apenas algumas sequências se encontram permanentemente disponíveis. Uma das ferramentas disponibilizada nesta base de dados é o *Software DrawNetwork* v.1.25. que permite a análise dos haplótipos de ADNmt, através da representação gráfica da estrutura genética das linhagens. Desta forma é possível detetar idiosincrasias e possíveis artefactos que possam resultar de problemas na sequenciação ou interpretação dos dados (Parson & Dür 2007).

3. Procedimento experimental

Na execução do procedimento experimental e consequente obtenção das sequências de ADNmt, foram utilizadas diversas técnicas laboratoriais recorrentes em genética e biologia molecular, implementadas no Serviço de Genética e Biologia Forenses do Instituto Nacional de Medicina Legal e Ciências Forenses, Delegação Sul (INMLCF-DS), que se encontram representadas de forma esquemática na Figura 9.

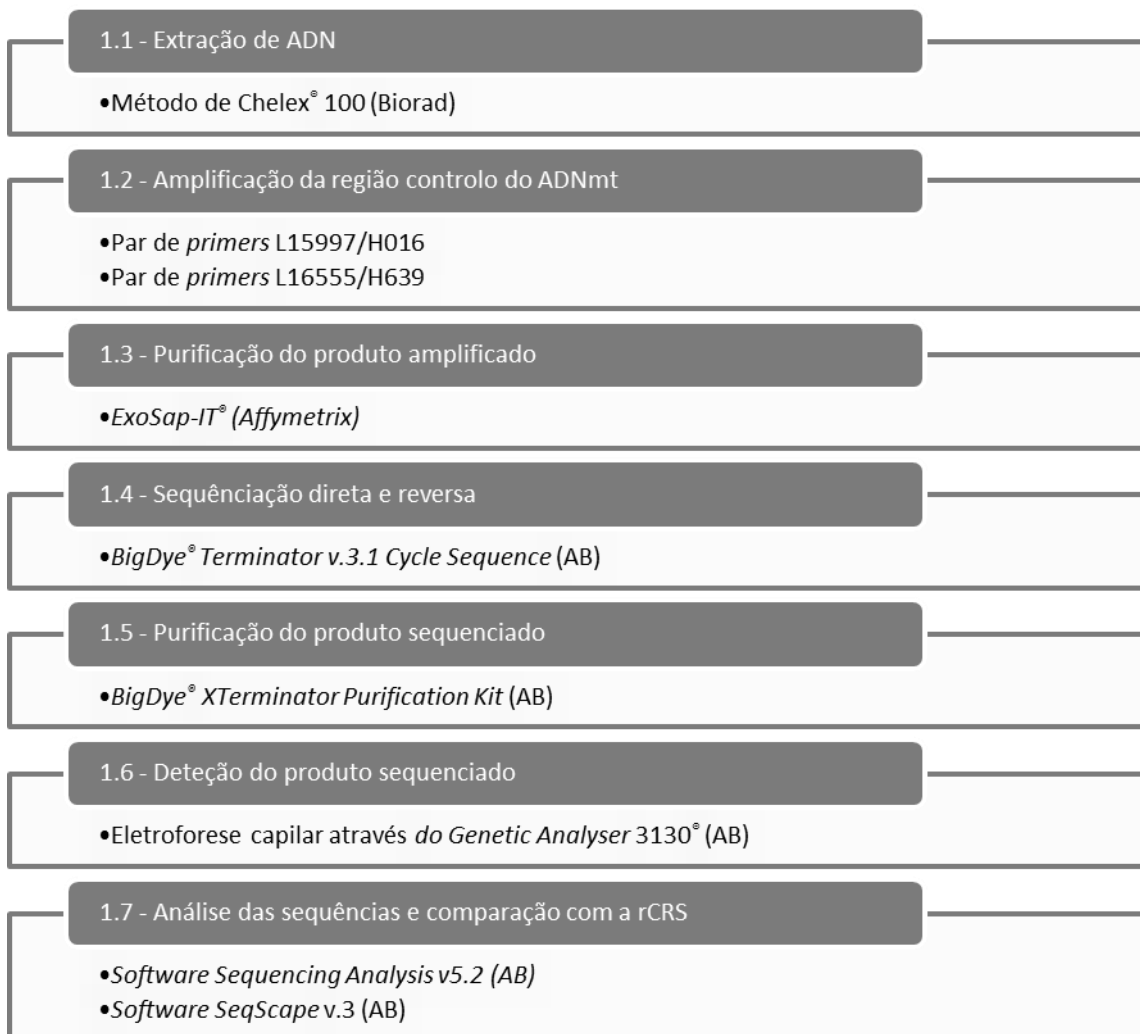


Figura 9 – Fluxograma representativo do procedimento experimental realizado para obtenção dos haplótipos de ADNmt.

3.1 Amostragem

No presente estudo foram utilizadas amostras de sangue de indivíduos de ascendência Angolana, residentes em Lisboa e não relacionados entre si, que compareceram no INMLC-DS, no âmbito de investigações de parentesco, entre os anos de 2003 e 2013. As amostras encontravam-se armazenadas no Serviço de Genética e Biologia Forenses e foram selecionados, de forma aleatória, indivíduos que preenchiem as características referidas anteriormente.

Foram estudadas manchas de sangue de 173 indivíduos, sendo 82 do sexo masculino e 91 do sexo feminino.

O conjunto de indivíduos selecionados poderá não ser representativo da população Angolana de origem, uma vez que não foi possível obter informação relativa à origem étnico-linguística dos indivíduos, mas sim representativo da população Angolana imigrante em Lisboa.

3.2 Extração de ADN

Para extração de ADN das amostras de sangue colhidas foi utilizada a resina quelante Chelex[®] 100 (BioRad) a 5%, de acordo com a técnica desenvolvida por Walsh e colaboradores em 1991 (Walsh *et al.* 1991). Esta resina é constituída por copolímeros de estireno-divinilbenzeno emparelhados com iões de iminodiacetato que têm a capacidade de atuar como agentes quelantes, estabelecendo ligação com iões metálicos polivalentes. Assim, os iões de magnésio presentes em meio biológico são atraídos pelo Chelex[®] 100; a remoção destes iões da solução leva à inativação da enzima nuclease responsável pela quebra das ligações entre os nucleótidos. Desta forma, fica assegurada a proteção das moléculas de ADN durante as várias fases do processo. Por outro lado, o Chelex[®] 100 é também responsável por agregar os restantes substratos não essenciais.

Numa fase prévia à extração foi necessária a preparação da solução de Chelex[®] 100 a 5% através da diluição 2,5 g da resina em 50mL de água desionizada. Esta preparação pode ser armazenada e utilizada em extrações posteriores, sendo que o seu acondicionamento não deve ultrapassar uma semana.

Na Figura 10 encontra-se esquematizado o processo de extração. Na prática, este processo iniciou-se com o corte de um círculo da mancha de sangue armazenada e conservada em papel FTA[®]. Este suporte contém químicos responsáveis pela lise celular e desnaturação proteica; tem também a capacidade de proteger os ácidos nucleicos de oxidação e de possível dano provocado por radiação UV. Por estes motivos o papel FTA[®] é frequentemente utilizado em armazenamento de longa duração de material biológico, podendo também ser utilizado no acondicionamento de plantas, insetos ou bactérias. O recorte, efetuado com um *paper-punch* de diâmetro 3mm (de forma a garantir a uniformização do tamanho das amostras extraídas) foi transferido para um tubo de microcentrífuga de 1,5mL

devidamente identificado com o código alfa numérico atribuído à amostra (Figura 10, ponto 1). A este recorte foi adicionado 1mL de água desionizada. Seguiu-se um período de incubação de 15min. à temperatura ambiente para que ocorresse a transferência do conteúdo do suporte para o solvente (Figura 10, ponto 2). Após uma ligeira agitação e centrifugação da mistura à velocidade de 14 000rpm (rotações por minuto) durante 3min., o material genético ficou concentrado no fundo do tubo de microcentrífuga, de onde se retiraram 970µL de sobrenadante (Figura 10, ponto 3 e ponto 4). Ao volume restante foram acrescentados 180µL da solução previamente preparada com a resina (Figura 10, ponto 5), após a qual ocorreu um novo período de incubação a 56⁰C durante 15min., garantindo a aderência dos resíduos ao Chelex[®] 100 (Figura 10, ponto 6). Para que as partículas da resina fossem pipetadas corretamente estas tiveram que ser mantidas em suspensão através de agitação (manual ou automática) e foram utilizadas pontas de micropipetas com um pequeno corte na extremidade para garantir a sucção das partículas. Na fase seguinte, foi efetuada uma nova agitação. De seguida, a exposição à temperatura de 100⁰C durante 8min. traduziu-se na desnaturação do ADN assim como na destruição de membranas e proteínas celulares (Figura 10, ponto 7). No último passo, uma centrifugação de 5min. a 14 000rpm agregou os resíduos celulares na parte inferior do tubo, tendo sido o sobrenadante utilizado na fase seguinte de amplificação. Em cada fase de extração foi efetuado um controlo negativo para o qual foi realizado todo o procedimento experimental anteriormente descrito, à exceção da adição da mancha de sangue.

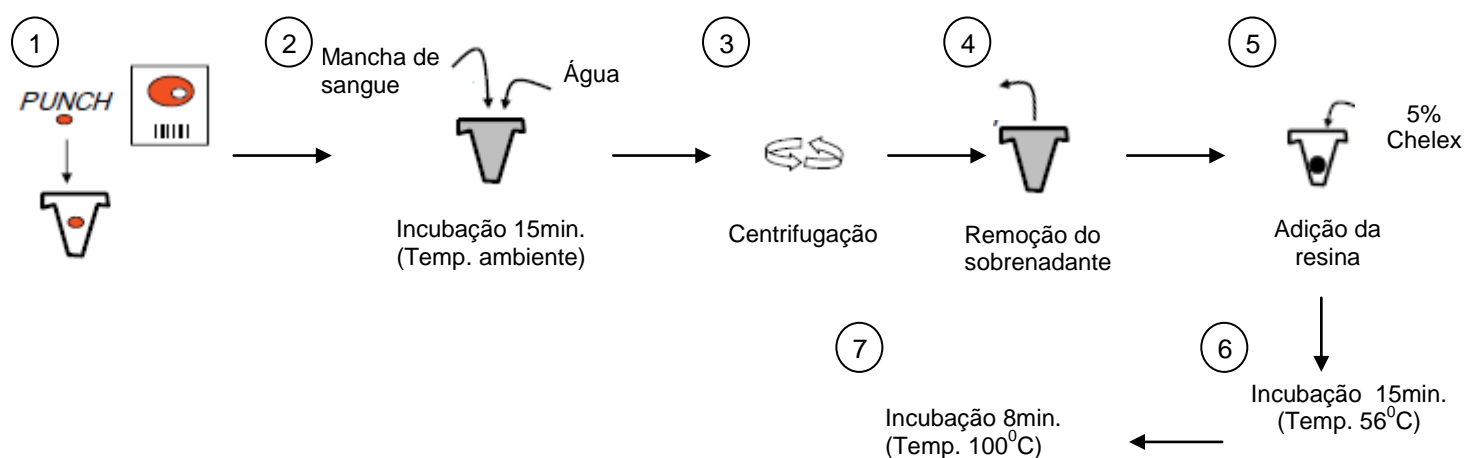


Figura 10 – Esquema do processo de extração de ADN através do método de Chelex[®] 100.
(Adaptado de Butler 2005).

3.3 Amplificação

A amplificação é uma técnica enzimática *in vitro* que se traduz na replicação de uma determinada região do ADN delimitada por *primers* oligonucleotídicos complementares das extremidades 3' de cada uma das sequências polinucleotídicas e tem como objetivo a

multiplicação dessas mesmas sequências (Erich 1989). Ao processo desenvolvido por Kary Mullis nos anos 80 foi dado o nome de *Polymerase Chain Reaction* (PCR) onde são diferenciadas três fases (K *et al.* 1986). Na primeira, denominada por desnaturação, as cadeias complementares da molécula de ADN ligadas por pontes de hidrogénio são separadas. De seguida os *primers* estabelecem ligação com as extremidades 3' de cada uma das cadeias polinucleotídicas, delimitando o fragmento a ser amplificado – fase de *annealing*. Por último, a enzima ADN polimerase sintetiza novas cadeias de ADN complementares à sequência alvo no processo denominado por extensão. Estas fases repetem-se por um determinado número de ciclos, podendo este variar, de forma a garantir o aumento exponencial do fragmento de ADN amplificado. Cada um destes passos requer uma temperatura ótima de atuação. Para que haja a separação das cadeias polinucleotídicas as moléculas têm que ser expostas à temperatura de cerca de 94°C. A descida de temperatura que se segue até cerca dos 60°C permite a ligação dos *primers*. Por último, pelos 72°C ocorre a atuação de ADN polimerase, realizando-se assim, a síntese das novas cadeias. Para além destas fases, é incluída ainda uma de temperatura de 94°C apenas no início do primeiro ciclo. No final do processo as amostras são mantidas e armazenadas a 4°C.

No caso particular do ADN mitocondrial foram utilizados dois pares de *primers* que garantiram a amplificação da região controlo total incluindo, desta forma, as regiões HVI, HVII e HVIII. O par L15997/H016 permitiu a amplificação entre as posições 16024 e 16569, enquanto o par L16555/H639 permitiu a amplificação entre as posições 1 e 576. Os *primers* são assim denominados pela cadeia com a qual interagem – L correspondente à cadeia leve (*Light*) e H à cadeia pesada (*Heavy*) – e pela posição nucleotídica na extremidade 3' (Tabela 2).

Tabela 2 - Sequência dos *primers* utilizados na amplificação.

L15997	5' CAC CAT TAG CAC CCA AAG CT 3'
H016	5' TGA TAG ACC TGT GAT CCA TCG TGA 3'
L16555	5' CCC ACA CGT TCC CCT TAA AT 3'
H639	5' GGG TGA TGT GAG CCC GTC TA 3'

O uso de dois *primers* para cada região ou conjunto de regiões permite a verificação e correção de erros que possam ocorrer e garante o esclarecimento de dúvidas associadas a um determinado nucleótido ou conjunto de nucleótidos (Parson *et al.* 2014).

Os principais componentes utilizados neste procedimento foram água desionizada, *Multiplex PCR Master Mix* (que contém *HotStarTaq Plus DNA Polymerase*, *PCR Buffer* com $MgCl_2$ numa concentração de 3mM, e dNTP's - desoxinucleótidos trifosfatados - a uma concentração de 400 μ M cada) (Qiagen), o conjunto de *primers* referidos anteriormente a uma concentração de 2 μ M e o ADN extraído na fase anterior. Os volumes necessários à reação estão descritos na Tabela 3.

Tabela 3- Volumes utilizados na preparação da mistura de reagentes necessários para amplificação da região controle total do ADNmt.

Reagente	Volume
<i>Multiplex PCR Master Mix</i>	5µL
<i>Primer</i>	1µL
H ₂ O	3µL
ADN	1µL
Total	10µL

A mistura foi efetuada para um total de n+1 amostras (sendo “n” o número de produtos a amplificar). Foram distribuídos 9µL da mistura de amplificação por cada tubo de amplificação de 0,5mL, aos quais foi posteriormente adicionado 1µL de ADN. Foi também preparado um controlo negativo para cada região a amplificar, onde se adicionou 1µL de água, em detrimento do ADN, para garantir a qualidade dos reagentes e averiguar a presença de contaminantes.

A reação de amplificação da região controlo total do ADNmt foi realizada num termociclador *GeneAmp PCR system 9700 (AB)* programado para as diferentes gamas de temperaturas representadas na Figura 11 e um total de 35 ciclos.

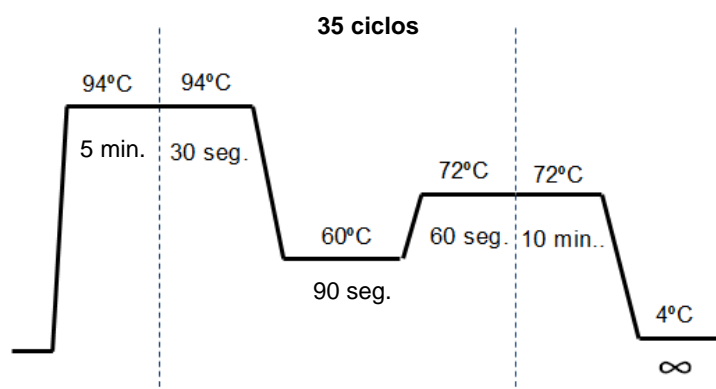


Figura 11– Representação das gamas de temperatura durante a *polymerase chain reaction* e intervalos de tempo correspondente (adaptado de Butler 2005).

3.4 Purificação dos produtos amplificados

Para a remoção de possíveis excessos de reagentes e seus componentes, resultantes do processo de amplificação, realizou-se a purificação do produto amplificado com o reagente *ExoSAP-IT (Affymetrix)*. Da sua constituição fazem parte duas enzimas hidrolíticas - exonuclease I e fosfatase alcalina recombinante de camarão. A primeira é responsável pela remoção de *primers* residuais enquanto a segunda garante a eliminação de dNTP's não utilizados no processo de amplificação.

Para cada 5µL de ADN foram necessários 2µL de *ExoSAP-IT*. A atividade das enzimas presentes nesta mistura foi iniciada por um período de 15min. à temperatura de 37°C. A finalização do processo ocorreu com a inativação das enzimas por um aquecimento a 80°C, de 15min.

A purificação do produto amplificado foi realizada num termociclador *GeneAmp PCR system 9700 (AB)* onde foram programadas as gamas de temperaturas necessárias.

3.5 Sequenciação

A sequenciação tem como objetivo determinar a posição exata de cada nucleótido de um segmento de ADN previamente amplificado. O método de sequenciação de Sanger, descrito por Frederick Sanger e colaboradores em 1977 (Sanger & Nicklen 1977), tem por base a utilização de didesoxinucleótidos trifosfatados – ddNTPs – cuja diferença relativamente aos desoxinucleótidos trifosfatados – dNTPs – está na ausência de grupo hidroxilo na extremidade 3', o que impossibilita a ligação fosfodiéster de um novo nucleótido. Desta forma, quando a polimerase introduz um novo ddNTP na cadeia a ser sintetizada, o crescimento desta termina. Cada um dos ddNTPs encontra-se marcado com um corante fluorescente que permite a sua deteção e identificação durante a análise. O ddTTP (timina) é marcado a vermelho, o ddCTP (citosina) a azul, ddATP (adenina) a verde e o ddGTP (guanina) a amarelo, embora seja exibido a preto, de forma a facilitar a visualização dos resultados obtidos (Figura 12).

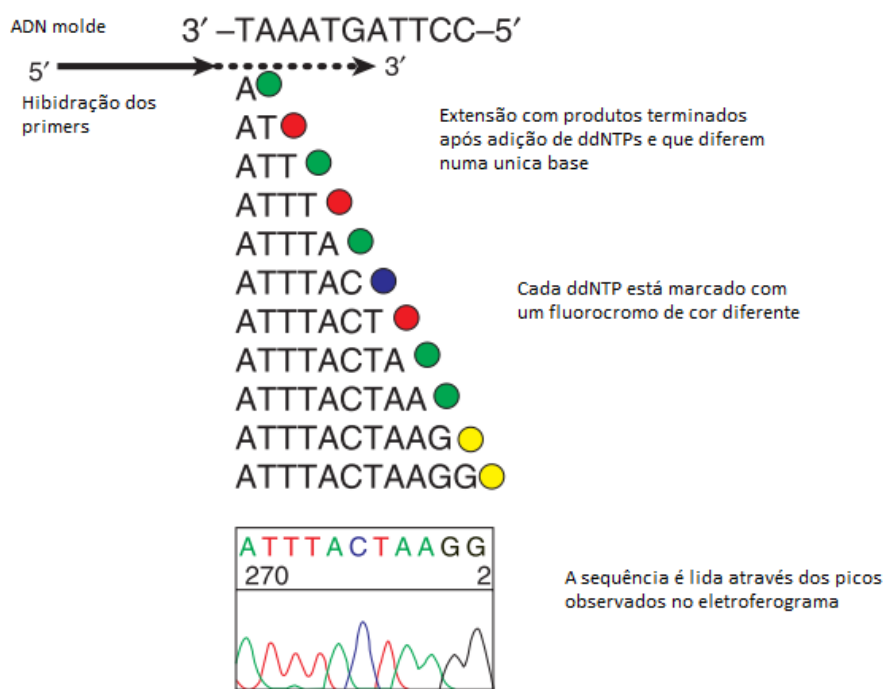


Figura 12 – Sequenciação de ADN com ddNTPs marcados com corante fluorescente (adaptado de Butler 2005).

Na reação necessária à sequenciação da região controlo total do ADNmt, para além dos *primers* referidos no ponto 3.3 (L15597, H016, L16555 e H639) a uma concentração de 2,5µM, foi também utilizada a química *BigDye Terminator v.3.1 Cycle Sequencing (AB)* e o tampão *Better buffer* (Microzone). Da química *BigDye Terminator v.3.1 Cycle Sequencing* fazem parte os ddNTPs (ddATP, ddCTP, ddGTP, ddTTP) marcados com *Dye Terminator*, os dNTPs (dATP, dCTP, dGTP, dTTP) e a ADN polimerase. O *Better buffer* foi responsável pela otimização da ação do *BigDye*, não sendo a sua composição química divulgada pelo fabricante. Os volumes necessários na reação de sequenciação estão descritos na Tabela 4 tendo sido preparada uma mistura para n+1 amostras, incluindo os negativos de amplificação. Foram ainda preparados negativos associados à fase de sequenciação, com a adição de 1µL de H₂O à mistura, em vez de produto amplificado. Nesta fase foi necessária a preparação de quatro reações, uma para cada *primer* de amplificação.

Tabela 4- Volumes utilizados na preparação da mistura de reagentes necessários para sequenciação da região controlo total do ADNmt.

Reagente	Volume
<i>DigDye Terminador</i>	2µL
<i>Better Buffer</i>	4µL
<i>Primer</i>	1µL
ADN	1µL
Total	8µL

O processo de sequenciação foi realizado num termociclador *GeneAmp PCR system 9700 (AB)* com programação prévia das gamas de temperatura e tempos representados na Figura 13. Esta fase do processo foi efetuada em placa de PCR de 96 poços com proteção de película aderente.

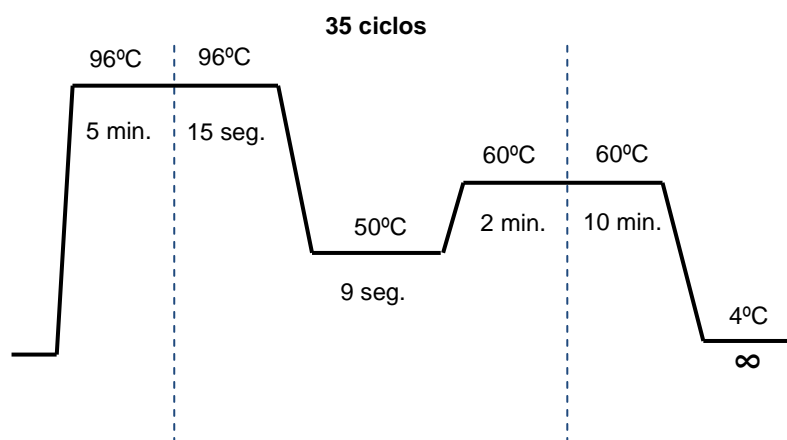


Figura 13 - Representação das gamas de temperatura durante a sequenciação, e intervalo de tempo correspondente (adaptado de Butler 2005).

3.6 Purificação dos produtos sequenciados

Da fase de sequenciação resultam excessos, sendo necessária uma nova purificação para que os dNTP'S e ddNTP's não emparelhados durante a sequenciação fossem removidos e não interferissem com a análise das sequências obtidas. Este processo foi efetuado com a ajuda do *BigDye Xterminator Purification Kit (AB)* que contem dois reagentes: *Xterminator solution* e *SAM TM solution*. O primeiro captura os dNTPs, ddNTPs e *primers* não incorporados e sais que se encontrem livres após o processo de sequenciação; o segundo melhora a atuação do reagente *Xterminator* e estabiliza a amostra após a fase de purificação.

Foi preparada uma mistura de reagentes de acordo com a Tabela 5, neste caso para um número de $n + q$ amostras, sendo “q” o número de corridas eletroforeticas realizadas, tendo em conta o conjunto de amostras preparadas na placa de 96 poços.

Tabela 5 – Volumes por reação necessários no processo de purificação dos produtos sequenciados.

Reagente	Volume
H ₂ O	30μL
SAM Solution	16,4μL
X Terminator	3,7μL
Total	50,1 μL

Distribuíram-se 50μL de amostra de reagentes por cada poço na placa de 96 poços, onde se encontravam distribuídas as amostras sequenciadas. O produto sequenciado juntamente com os reagentes de purificação foram colocados em agitação durante 30 min., período durante o qual ocorreu a adesão e a captura dos resíduos não desejados pelas partículas do *Xterminator*. Seguiu-se uma centrifugação de 2 min a 2800rpm que permitiu a concentração dos reagentes de purificação e dos constituintes agregados no fundo da placa, impedindo que estes fossem aspirados pelos capilares durante a eletroforese capilar.

3.7 Detecção dos produtos sequenciados

A detecção dos fragmentos obtidos durante a fase de sequenciação e aquisição da sequência de ADNmt foi realizada por eletroforese capilar num sequenciador *Genetic Analyser 3130 (AB)* com 4 capilares (1 capilar por amostra) (Figura 14). Os diferentes fragmentos migraram de acordo com o seu peso molecular, sendo que os de menor peso molecular percorrem os capilares a uma maior velocidade que os de maior peso molecular. A detecção foi feita pela presença dos ddNTPs marcados com fluorocromos no final da cadeia, que ao serem excitados por um feixe de laser incidente no capilar, emitiram fluorescência. Esta foi representada por um pico, de cor correspondente ao ddNTP excitado, no eletroferograma.

Na eletroforese capilar podem ser utilizados, como matriz de separação, os polímeros POP-6 (AB) e POP-7 (AB), que ao revestirem o capilar, permitem controlar a corrente elétrico-osmótica. O POP-6 é considerado o mais adequado no estudo de ADNmt devido à constituição da malha, que garante uma separação mais eficiente.

O sequenciador *Genetic Analyser 3130* (AB) contém cinco reservatórios cujo conteúdo deve ser substituído previamente à sua utilização. Três deles contêm água e os restantes dois tampão (um correspondente ao ânodo e outro ao cátodo), cada um deles com aproximadamente 16mL da respetiva solução. A preparação do tampão foi realizada com a diluição de 3,5mL de tampão EDTA (10x, AB) em 31,5mL de água desionizada; desta forma, metade do volume foi revertida para o reservatório correspondente ao ânodo da eletroforese e a restante para o cátodo (Figura 15).



Figura 14 – Sequenciador Genetic Analyser 3130 (AB) do serviço de Genética e Biologia Forenses do INMLCF-DS.



Figura 15– Genetic Analyser 3130 (AB), vista interior. (1) Reservatório ânodo; (2) Reservatório cátodo.

Para preparar a corrida eletroforética utilizou-se o *software Data Collection v.4* que permitiu a identificação das amostras presentes em cada poço da placa de 96 poços.

3.8 Análise dos produtos sequenciados

Como resultado da deteção dos produtos amplificados foram obtidos, para uma mesma amostra, dois segmentos com sequências que perfazem a totalidade da região controlo do ADNmt. Consequentemente, para cada um deles, foram apresentadas duas sequências, correspondentes à amplificação obtida pelo *primer* associado à cadeia leve (L15997 ou L16555) e pelo *primer* associado à cadeia pesada (H016 ou H639).

Na análise dos resultados do ADNmt foram utilizados dois *softwares*: *Sequencing Analysis v.5.2* e *SeqScape V.3* (AB). O primeiro permitiu avaliar a qualidade da sequência através da identificação dos nucleótidos em cada posição, que é representado por um pico no eletroferograma de cor correspondente ao fluorocromo associado. Na Figura 16 é apresentado um exemplo de um segmento analisado no *software Sequencing Analysis*.

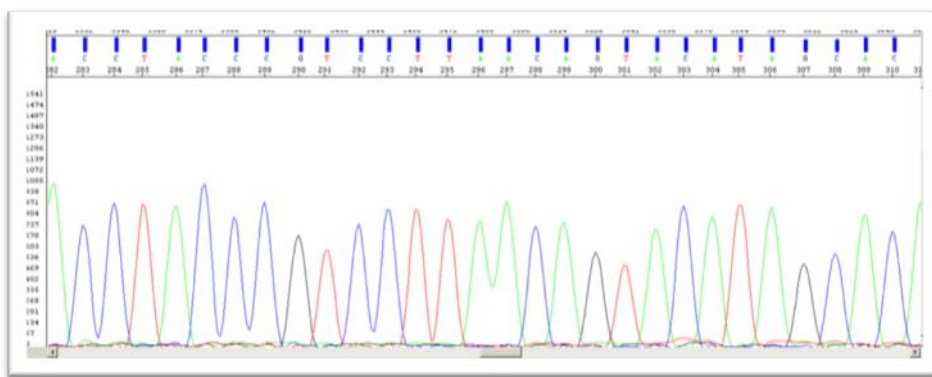


Figura 16 – Segmento de uma sequência de ADNmt obtido pelo *primer* L16555, analisada pelo *software* *Sequencing Analysis* v.5.2.

Cada pico no eletroferograma corresponde a um nucleótido dessa sequência. A barra na parte superior de cada posição (A,T,G e C) representa a qualidade e individualização do nucleótido dada pelo *software*. Uma barra vermelha representará um pico mal definido onde não é possível identificar o nucleótido presente nessa posição; uma barra amarela significa que, apesar de se conseguir identificar o nucleótido, existem dúvidas que devem ser esclarecidas; uma barra azul representa um pico corretamente identificado e para o qual não existe qualquer dúvida (Figura 17). As incongruências associadas à presença de barras amarelas e vermelhas revelam ser mais comuns no início e no final das sequências.

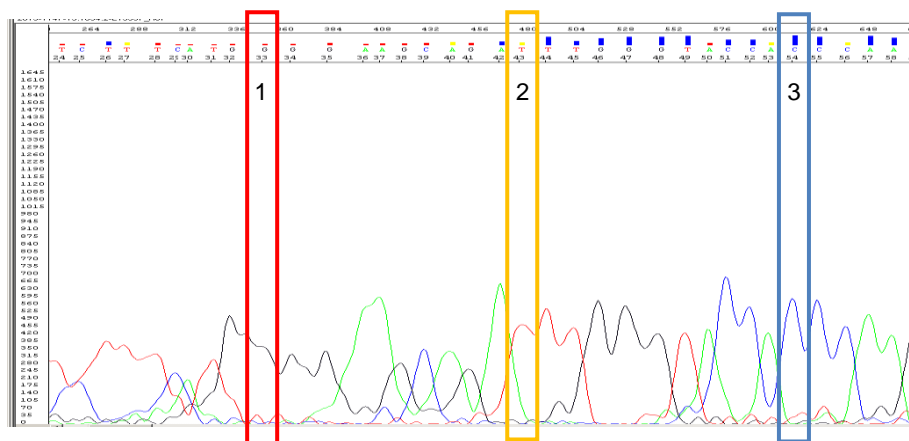


Figura 17 - Eletroferograma obtido pelo *primer* L15997. (1) Nucleótido que origina muitas dúvidas; (2) Nucleótido suscetível a dúvidas; (3) Nucleótidos corretamente identificado.

Após a verificação da qualidade das sequências obtidas estas foram posteriormente analisadas no *software* *SeqScape* v.3 que permitiu o alinhamento e a comparação das sequências em estudo com a rCRS e com a identificação dos polimorfismos existentes nas sequências estudadas. O conjunto destes polimorfismos característicos de uma sequência de ADNmt pertencente a um determinado indivíduo dá origem ao haplótipo. A sua caracterização deve ser realizada tendo por base regras padronizadas internacionalmente por forma a garantir uma uniformização no estudo do ADNmt. Neste sentido, seguiram-se as diretrizes descritas

pelo *European DNA Profiling Group* (EDNAP) (Tully *et al.* 2001) e as regras de nomenclatura de acordo com as normas da International Union of Pure and Applied Chemistry (IUPAC) (Tabela 6).

Tabela 6 – Código IUPAC utilizado na análise do ADNmt.

Código IUPAC	Base(s) nucleotídica(s)	Código IUPAC	Base(s) nucleotídica(s)
A	Adenina	Y	Citosina ou Timina (pYrimidine)
C	Citosina	K	Guanina ou Timina (<i>Keto</i>)
G	Guanina	M	Adenina ou Citosina (<i>aMino</i>)
T	Timina	N	aNy base
B	Citosina, Guanina ou Timina	V	Adenina, Citosina ou Guanina
D	Adenina, Guanina ou Timina	S	Guanina ou Citosina (<i>Strong bonds</i>)
H	Adenina, Citosina ou Timina	W	Adenina ou Timina (<i>Weak bonds</i>)
R	Adenina ou Guanina (<i>puRine</i>)		

A referência a substituições nas sequências de ADNmt em estudo foi realizada relativamente à rCRS, com a identificação da posição em que a alteração ocorreu e do nucleótido que originou essa mesma alteração. Esta classificação é realizada de acordo com a terminologia criada pela IUPAC, em letras maiúsculas. Como exemplo tem-se o polimorfismo existente na posição 73 do segmento HVII da região controlo do ADNmt. Na rCRS esta posição é caracterizada por uma adenina (A), contudo em grande parte dos haplótipos verificou-se a substituição por uma guanina (G), sendo o polimorfismo identificado como **73G** (Carracedo *et al.* 2000; Parson *et al.* 2014; Scientific Working Group for DNA Analysis Methods 2003).

As inserções foram identificadas pela posição nucleotídica anterior à qual ocorreu a alteração, seguida de .1 na primeira inserção, .2 na segunda inserção, e assim sucessivamente; este número foi ainda seguido pelo nucleótido inserido. A inserção de uma Adenina (A) característica da posição 524 é classificada como **524.1A**.

As deleções foram representadas pela posição do nucleótido que foi deletado juntamente com o sufixo “DEL”, “del” ou “-”. Um polimorfismo deste tipo é exemplificado pela remoção de uma Adenina (A) na posição 523 sendo identificada como **523DEL** ou **523del** (Parson *et al.* 2014; Carracedo *et al.* 2000).

No caso particular de inserções ou deleções em regiões homopoliméricas, a localização exata da inserção é desconhecida; desta forma o correto será identificá-las na extremidade 3' dessa mesma região. A título elucidativo, tem-se a inserção de uma Citosina (C) comumente visualizada entre as posições 311 e 315. Esta região contém 5 Citosinas (C) e é relativamente polimórfica; a alteração é representada por uma inserção identificada na última posição desta região sendo descrita como **315.1C** (Carracedo *et al.* 2000; Scientific Working Group for DNA Analysis Methods 2003; Parson *et al.* 2014).

Em amostras onde se verificou a presença de heteroplasmia de comprimento, a leitura da sequência obtida pelo *primer* direto ou pelo *primer* reverso, apenas foi possível até à posição heteroplasmica. A partir desta posição, o eletroferograma era composto por uma mistura de bases nucleotídica. Nestes casos, foi necessário realizar uma repetição da amplificação do segmento heteroplasmico para que fosse possível confirmar e validar a sequência (Figura 18) (Parson *et al.* 2014).

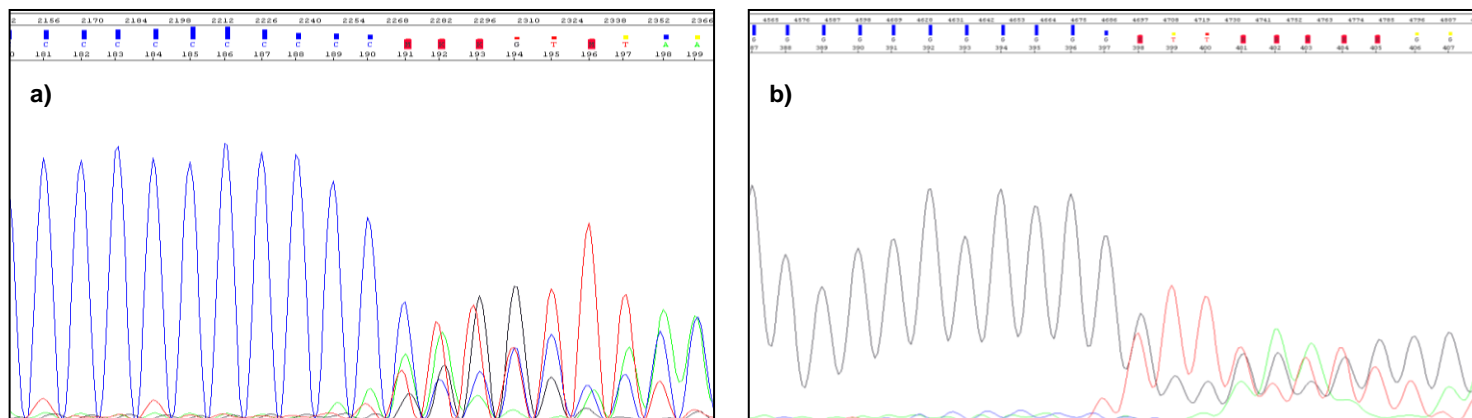


Figura 18 – Eletroferogramas de uma amostra com heteroplasmia de comprimento numa cadeia de ADNmt da região HVII, analisado pelo *software Sequencing Analysis v.5.2*. a) sequenciação com o *primer* associado à cadeia leve - L16555; b) sequenciação com o *primer* associado à cadeia pesada - H639.

A heteroplasmia de posição é representada pela sobreposição de nucleótidos numa mesma posição da sequência de ADNmt o que se traduz em dois picos no eletroferograma; a altura conjunta dos dois será aproximada da altura dos picos que rodeiam essa posição heteroplasmica (Figura 19). A identificação da heteroplasmia é realizada pela posição em que ocorre seguida da letra correspondente, de acordo com o código descrito pela IUPAC. Em casos em que a posição heteroplasmica resulta de inserções ou deleções, a classificação deve ser realizada com letras minúsculas (Parson *et al.* 2014).

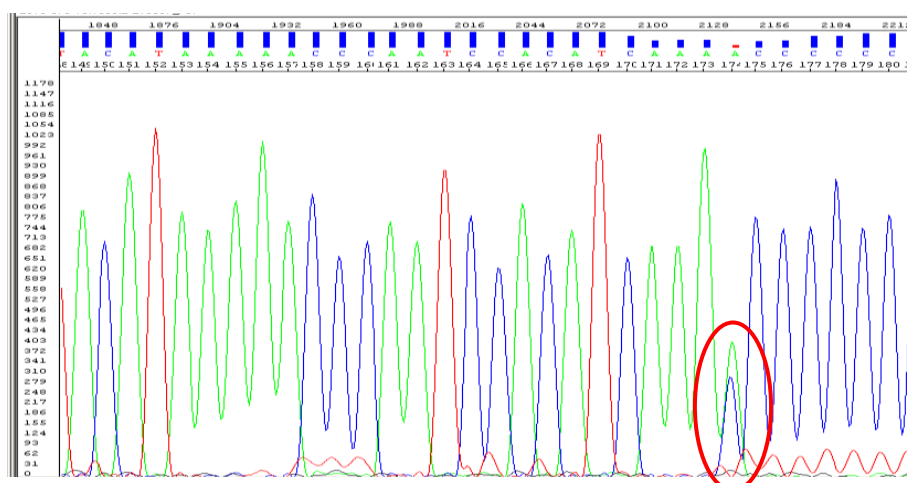


Figura 19– Heteroplasmia de posição identificada a vermelho.

A atribuição de haplogrupos a cada um dos haplótipos estudados foi efetuada com recurso à *PhyloTree*, árvore filogenética de ADNmt desenvolvida por van Oven e Kayser em 2009 (Figura 20) (van Oven & Kayser 2009; Bandelt *et al.* 2012). A *PhyloTree* não considera para construção da árvore filogenética de ADNmt as mutações 309.1C(C), 315.1C, 16182C, 16183C, 16193.1C(C), 16519C e as inserções/deleções nas posições 515-522.

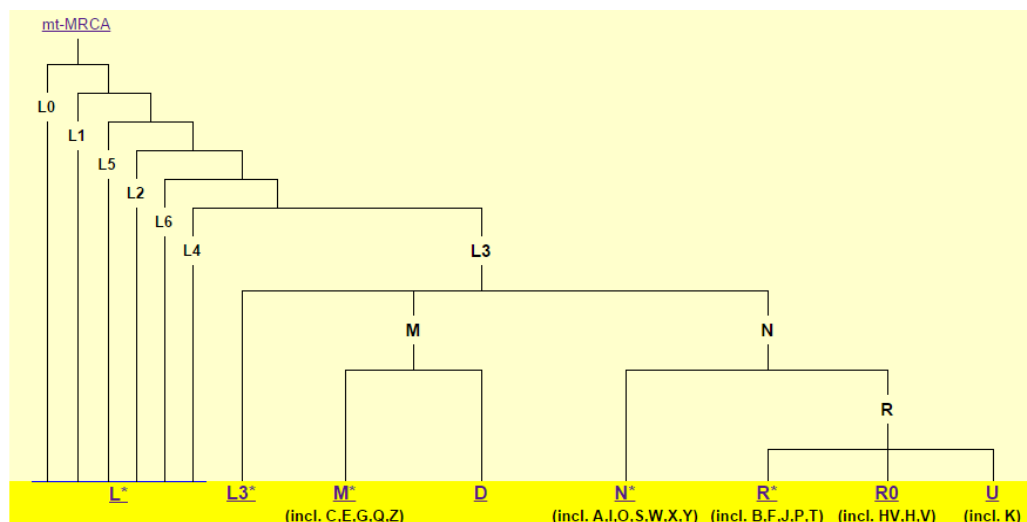


Figura 20 – Árvore filogenética de ADNmt (van Oven & Kayser 2009)

3.9 Análise estatística e filogenética

Os valores de diversidade nucleotídica e diversidade de sequência referentes às 173 sequências de ADNmt de imigrantes Angolanos a residir em Lisboa, foram obtidos para três segmentos distintos: para a totalidade da região controlo do ADNmt, para as regiões HVI e HVII da região controlo do ADNmt e para a região HVI isolada da região controlo do ADNmt. Estes cálculos foram realizados manualmente através da metodologia desenvolvida por Nei e Tajima em 1981 e pelo *software* Arlequin (Excoffier *et al.* 2005).

A análise de variância molecular (AMOVA) determina a estrutura genética das populações. Este parâmetro permite calcular a variabilidade genética entre populações através do estudo dos haplótipos e das suas frequências. Para efetuar a AMOVA e, consequentemente, determinar as distâncias genéticas interpopulacionais, recorreu-se ao programa Arlequin v. 3.5.1.2., através do qual foram obtidos os valores de F_{st} (*fixation index*) e *p-value* para a região controlo total do ADNmt e para as regiões HVI e HVII. O parâmetro F_{st} permite o cálculo de diferenciação entre populações atribuído com base na sua estrutura genética e varia entre 0 e 1. O *p-value* é analisado comparativamente ao nível de significância (0.05). As distâncias moleculares entre populações foram obtidas através do programa *Neighbor* do *software* *Phylip* v. 3.69 (Felsenstein 1989) e com base nos valores F_{st} calculados. A representação gráfica em forma de árvore filogenética foi realizada com o *software* *TreView* v. 1.5.2 (Page 1996).

4. Resultados

No presente estudo foi sequenciada e analisada a totalidade da região controlo do ADNmt de amostras sanguíneas de 173 indivíduos de ascendência Angolana, não relacionados entre si e residentes em Lisboa. A amostragem inicial continha 176 indivíduos, contudo 3 deles não foram incluídos no estudo, uma vez que as características particulares das amostras colhidas impediram a obtenção de resultados com a qualidade desejada.

Como principal objetivo do trabalho pretendeu-se determinar os haplótipos e respetivos haplogrupos de ADNmt dos indivíduos Angolanos a residir em Lisboa. Desta forma será possível obter a estrutura genética de ADNmt da população em estudo e avaliar o impacto destes indivíduos imigrantes na região de Lisboa.

O haplótipo é o conjunto de polimorfismos identificados nas respetivas posições nucleotídicas por comparação com a rCRS e de acordo com as diretrizes descritas para a classificação de sequências de ADNmt (Tully *et al.* 2001; Parson *et al.* 2014). A discriminação dos haplogrupos foi realizada recorrendo à árvore filogenética desenvolvida por van Oven e Kayser em 2009 (van Oven & Kayser 2009), versão 16. Na Tabela 7 estão representados os polimorfismos que formam os haplótipos assim como a identificação dos respetivos haplogrupos em que se inserem.

Os haplótipos descritos neste estudo foram submetidos à base de dados EMPOP (Parson & Dür 2007) e aceites para inserção na mesma. Os haplótipos encontram-se atualmente disponíveis com o código de acesso EMPOP662.

Tabela 7 - Identificação do código da sequência, respetivo haplogrupo e haplótipo com os respetivos polimorfismos referentes à rCRS.

Amostra	Haplogrupo	Haplótipo
FS_ANG002	K2a1a	16066G 16182C 16183C 16189C 16223T 16224C 16278T 16390A 16519C 73G 146C 152C 182T 263G 315.1C 511T
FS_ANG003	L2a1g	16131C 16189C 16223T 16225T 16234T 16278T 16294T 16309G 16390A 73G 146C 152C 195C 263G 309.1C 309.2C 315.1C
FS_ANG004	L3e2b	16172C 16182C 16183C 16189C 16223T 16320T 16519C 73G 150T 195C 263G 315.1C
FS_ANG005	L2a1	16092C 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
FS_ANG006	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 16390A 89C 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG007	L3e2b	16172C 16183C 16189C 16223T 16320T 16519C 73G 150T 195C 263G 315.1C
FS_ANG008	L1c1	16038G 16129A 16187T 16189C 16223T 16271C 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 521G 523DEL 524DEL
FS_ANG009	L1c1b	16086C 16129A 16187T 16189C 16223T 16241G 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 416A 518T 523DEL 524DEL
FS_ANG010	L1c3b1a	16129A 16163G 16187T 16189C 16209C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523DEL 524DEL
FS_ANG011	H	16519C 152C 263G 309.1C 315.1C
FS_ANG012	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG013	L2a1a3c	16093C 16223T 16256T 16278T 16294T 16309G 16390A 16519C 73G 143A 146C 152C 195C 263G 315.1C
FS_ANG016	L4b2	16093G 16223T 16287A 16293T 16301T 16311C 16355T 16362C 16399G 73G 146C 150T 152C 195C 200G 244G 263G 315.1C 513A
FS_ANG018	L1c3a	16129A 16182C 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16360T 16519C 73G 152C 182T 186A 189C 247A 263G 315.1C 316A 523DEL 524DEL
FS_ANG017	L0a2a2	16148T 16172C 16173T 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG019	L1c1b	16086C 16129A 16187T 16189C 16223T 16241G 16278T 16291T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 416A
FS_ANG020	L3e1	16176T 16223T 16327T 73G 150T 189G 200G 263G 315.1C
FS_ANG021	L1b1a10	16126C 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 151T 152C 182T 185T 195C 247A 263G 315.1C 357G 523DEL 524DEL 567G
FS_ANG022	L3e2b	16164G 16172C 16189C 16223T 16320T 16519C 73G 146C 150T 195C 263G 315.1C 523DEL 524DEL
FS_ANG023	L3e1	16207G 16223T 16327T 73G 150T 183G 189G 200G 263G 315.1C
FS_ANG024	L3f2	16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 315.1C
FS_ANG025	L3d	16124C 16223T 73G 152C 207A 263G 315.1C 523DEL 524DEL
FS_ANG026	L3e1	16213A 16223T 16309G 16327T 73G 150T 189G 200G 263G 315.1C
FS_ANG027	T1a9	16126C 16163G 16186T 16189C 16249C 16294T 16311C 16519C 73G 263G 309.1C 315.1C
FS_ANG028	L1c1	16129A 16187T 16189C 16223T 16263C 16278T 16293G 16294T 16311C 16360T 16368C 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG029	L3b1a	16124C 16223T 16278T 16362C 16519C 73G 152C 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG030	L1b	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523DEL 524DEL
FS_ANG031	T2c1	16126C 16292T 16294T 16296T 16519C 73G 146C 263G 315.1C 523DEL 524DEL
FS_ANG032	L3f2	16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 315.1C
FS_ANG033	R0	16519C 263G 315.1C
FS_ANG034	L2a1f3	16189C 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 198T 263G 315.1C
FS_ANG035	L2a1a	16223T 16265C 16278T 16294T 16309G 16368C 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C 573.1C 573.2C 573.3C 573.4C
FS_ANG036	L2b	16114A 16129A 16212G 16213A 16223T 16278T 16390A 73G 146C 150T 152C 182T 195C 198T 204C 207A 263G 309.1C 309.2C 315.1C
FS_ANG037	L3e4	16051G 16223T 16264T 16519C 73G 150T 257G 263G 315.1C 523DEL 524DEL
FS_ANG038	L1c1b	16086C 16129A 16187T 16189C 16223T 16241G 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 416A 518T 523DEL 524DEL
FS_ANG039	L0a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG040	L0a1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16234T 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG041	L1b1a10	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 151T 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523DEL 524DEL
FS_ANG042	L1c3c	16129A 16187T 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 73G 93G 151T 152C 182T 186A 189C 195C 247A 248G 263G 315.1C 316A 458T 523DEL 524DEL
FS_ANG043	J1c2c	16069T 16126C 16519C 73G 146C 185A 188G 228A 263G 295T 309.1C 315.1C 462T 489C
FS_ANG044	L3b	16124C 16223T 16278T 16362C 16519C 73G 263G 315.1C 523DEL 524DEL
FS_ANG045	L2c2b1b	16223T 16264T 16278T 16311C 16390A 73G 93G 146C 150T 152C 182T 183G 195C 198T 263G 264T 315.1C 325T 523DEL 524DEL
FS_ANG046	L3b1a3	16124C 16223T 16278T 16311C 16362C 16519C 73G 263G 315.1C 523DEL 524DEL
FS_ANG047	L2c2b1b	16223T 16264T 16278T 16311C 16390A 73G 93G 146C 150T 152C 182T 183G 195C 198T 263G 315.1C 325T 523DEL 524DEL
FS_ANG048	L1c1a2	16129A 16183C 16189C 16223T 16274A 16278T 16293G 16294T 16311C 16360T 16519C 73G 89C 93G 95C 152C 182T 186A 189C 236C 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG049	H1b	16356C 16519C 263G 315.1C
FS_ANG050	L1b	16126C 16187T 16189C 16223T 16264T 16270T 16274A 16278T 16293G 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 373G 523DEL 524DEL
FS_ANG051	L0a1b	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 152C 185A 189C 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG052	L2a1e1	16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C 524.1A 524.2C 524.3A 524.4C
FS_ANG053	L3e1a	16185T 16223T 16311C 16327T 73G 150T 185A 189G 200G 263G 315.1C
FS_ANG054	L1c1a	16093C 16129A 16187T 16189C 16223T 16263C 16278T 16293G 16294T 16311C 16360T 16368C 16519C 73G 151T 152C 182T 186A 189C 195C 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG055	L2a1g	16131C 16189C 16223T 16225T 16234T 16278T 16294T 16309G 16390A 73G 146C 152C 195C 263G 309.1C 309.2C 315.1C
FS_ANG056	L2b2a	16114A 16129A 16213A 16223T 16278T 16354T 16390A 73G 146C 150T 152C 182T 195C 198T 204C 263G 315.1C
FS_ANG057	L1c1	16038G 16187T 16189C 16223T 16261T 16278T 16293G 16294T 16311C 16360T 16368C 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG058	L2c2b1b	16223T 16264T 16278T 16311C 16390A 73G 93G 146C 150T 152C 182T 183G 195C 198T 263G 315.1C 325T 523DEL 524DEL
FS_ANG059	L3b1a8	16145A 16223T 16278T 16362C 16519C 73G 195C 263G 279C 315.1C 523DEL 524DEL
FS_ANG060	L0a2a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 93G 152C 189G 204C 207A 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG061	L0a2a2	16086C 16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG062	L1c	16129A 16183C 16189C 16223T 16278T 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523DEL 524DEL
FS_ANG063	L1c3b1a	16129A 16163G 16187T 16189C 16209C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 247A 263G 309.1C 315.1C 316A 523DEL 524DEL
FS_ANG064	L3d3a	16124C 16183C 16189C 16223T 16278T 16304C 16311C 73G 152C 263G 315.1C 523DEL 524DEL
FS_ANG065	H	16519C 152C 263G 309.1C 315.1C
FS_ANG066	L2a1c1	16086C 16223T 16278T 16294T 16309G 16390A 73G 143A 146C 152C 195C 198T 263G 315.1C
FS_ANG067	L0a'b	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16274A 16278T 16311C 16320T 89C 93G 95C 152C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG068	L2b	16114A 16129A 16213A 16223T 16274A 16278T 16293G 16519C 73G 146C 150T 152C 182T 183G 195C 198T 204C 263G 309.1C 315.1C
FS_ANG069	L2a1a1	16223T 16278T 16294T 16309G 16368C 16390A 16519C 73G 146C 195C 263G 309.1C 315.1C 524.1A 524.2C 524.3A 524.4C
FS_ANG070	L3d3a	16124C 16183C 16189C 16223T 16278T 16304C 16311C 73G 152C 263G 315.1C 523DEL 524DEL
FS_ANG071	L3b	16124C 16223T 16278T 16362C 16399G 16519C 73G 263G 315.1C 523DEL 524DEL
FS_ANG072	L3e2a1b1	16129A 16223T 16320T 16399G 16519C 73G 150T 195C 198T 263G 315.1C 499A
FS_ANG073	L0d1b2	16129A 16187T 16189C 16223T 16230G 16243C 16294T 16311C 16519C 73G 146C 152C 195C 199C 247A 315.1C 316A 498DEL
FS_ANG074	L1c3a1b	16129A 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16390A 16519C 73G 151T 152C 182T 186A 189C 247A 263G 309.1C 309.2C 315.1C 316A 523DEL 524DEL
FS_ANG075	L3d3a	16124C 16183C 16189C 16223T 16278T 16304C 16311C 73G 152C 263G 315.1C 523DEL 524DEL
FS_ANG076	L1b	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 309.1C 315.1C 357G 523DEL 524DEL
FS_ANG077	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG078	L3e1a	16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG079	L3f2	16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG080	L0a1ba	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 150T 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG081	L3e1d	16176T 16223T 16327T 73G 150T 152C 189G 200G 263G 315.1C
FS_ANG082	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16234T 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG083	K1a1b1c	16222T 16224C 16311C 16327T 16519C 73G 263G 315.1C 385G 497T
FS_ANG084	L1c1a2	16129A 16183C 16189C 16223T 16274A 16278T 16293G 16294T 16311C 16360T 16519C 73G 76T 89C 93G 95C 152C 182T 186A 189C 236C 247A 263G 297G 309.1C 315.1C 316A 523DEL 524DEL
FS_ANG085	L1c2b2	16078G 16129A 16187T 16189C 16223T 16265C 16286A 16294T 16311C 16320T 16360T 16519C 16527T 73G 151T 152C 186A 189C 195C 198T 204C 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG086	L3e1	16223T 16327T 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG087	L1c1	16093C 16129A 16187T 16189C 16274A 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 249DEL 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG088	L1c2b1b1	16093C 16129A 16187T 16189C 16223T 16265C 16278T 16286C 16294T 16311C 16360T 16519C 16527T 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 264T 297G 315.1C 316A
FS_ANG089	L2a1a2	16223T 16278T 16286T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
FS_ANG090	L3e3	16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523DEL 524DEL

Na tabela: A- Adenina, C- Citosina; G- Guanina; T- Timina; DEL- Deleção; M – Adenina e Citosina

Tabela 7 (continuação) - Identificação do código da sequência, respetivo haplogrupo e haplótipo com os respetivos polimorfismos referentes à rCRS.

Amostra	Haplogrupo	Haplótipo
FS_ANG091	U5a2a	16114A 16192T 16256T 16270T 16294T 16526A 73G 195C 263G 309.1C 315.1C
FS_ANG092	L3e2b	16172C 16189C 16223T 16311C 16320T 16519C 73G 150T 152C 195C 263G 315.1C
FS_ANG093	L1c1b	16086C 16129A 16187T 16189C 16223T 16241G 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A 416A 518T 523DEL 524DEL
FS_ANG094	R0	16519C 263G 315.1C
FS_ANG095	L3e2b	16172C 16183C 16189C 16223T 16320T 16519C 73G 150T 195C 263G 309.1C 315.1C
FS_ANG096	L2a1	16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 263G 309.1C 315.1C
FS_ANG097	L2a1	16189C 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
FS_ANG098	L0a2a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG099	U5a1	16256T 16270T 16399G 73G 263G 309.1C 309.2C 315.1C
FS_ANG100	L3e3b2	16223T 16265T 16519C 16527T 73G 150T 195C 263G 315.1C 523DEL 524DEL
FS_ANG101	L2a1a3	16129A 16223T 16278T 16294T 16309G 16390A 73G 143A 146C 152C 195C 263G 315.1C
FS_ANG102	L3d1a1a	16124C 16223T 16319A 73G 146C 150T 263G 315.1C 523DEL 524DEL
FS_ANG103	L3d1d	16124C 16223T 16256T 16368C 73G 152C 263G 315.1C 523DEL 524DEL
FS_ANG104	L1c3a	16129A 16182C 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16360T 16519C 73G 152C 182T 186A 189C 247A 263G 315.1C 316A 573.1C 573.2C 573.3C 573.4C
FS_ANG105	L3e2b	16172C 16189C 16223T 16320T 16366T 16519C 73G 150T 195C 263G 309.1C 315.1C 573.1C 573.2C
FS_ANG106	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG107	L0a1b1	16071T 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16261T 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG108	L0a2a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 95C 152C 189G 204C 207A 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG109	L1c3b1a	16129A 16163G 16187T 16189C 16209C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523DEL 524DEL
FS_ANG110	L2a1a2	16223T 16278T 16286T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
FS_ANG111	L3f1b1a1	16129A 16209C 16223T 16292T 16295T 16311C 16519C 73G 189G 200G 263G 272G 309.1C 315.1C
FS_ANG112	L2c2	16223T 16264T 16278T 16390A 73G 93G 146C 150T 152C 182T 195C 198T 263G 309.1C 315.1C 325T 523DEL 524DEL
FS_ANG113	L1c3b	16129A 16163G 16187T 16189C 16209C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 247A 263G 315.1C 316A 523DEL 524DEL
FS_ANG114	L2a1'2'3'4	16223T 16278T 16293G 16294T 16390A 73G 143A 146C 152C 182T 195C 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG115	L1c3b'c	16129A 16187T 16189C 16223T 16274A 16278T 16293G 16294T 16311C 16360T 16519C 73G 93G 95C 151T 152C 154C 182T 186A 189C 236C 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG116	L0a2a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG117	L3e2b	16172C 16182C 16183C 16189C 16223T 16320T 16519C 73G 150T 152C 195C 263G 315.1C
FS_ANG118	L1c2	16092C 16129A 16187T 16189C 16223T 16265C 16286G 16294T 16311C 16360T 16519C 16527T 73G 151T 152C 182T 186A 189C 195C 198T 200G 247A 263G 297G 315.1C 316A
FS_ANG119	L3e1	16223T 16327T 73G 150T 189G 200G 263G 315.1C
FS_ANG120	U5a2a	16114A 16192T 16256T 16270T 16294T 16526A 73G 195C 263G 309.1C 315.1C
FS_ANG121	L1c1a	16129A 16182C 16183C 16189C 16223T 16278T 16293G 16294T 16311C 16355T 16360T 16519C 73G 89C 93G 95C 151T 152C 182T 186A 189C 195C 236C 247A 263G 297G 315.1C 316A 523DEL
FS_ANG122	L1b	16126C 16187T 16189C 16223T 16264T 16270T 16278T 16311C 16519C 73G 152C 182T 185T 195C 247A 263G 315.1C 357G 523DEL 524DEL
FS_ANG123	L0a1a2	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 64T 93G 95C 151T 185A 189G 200G 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG124	L3d	16124C 16223T 73G 152C 207A 263G 315.1C 523DEL 524DEL
FS_ANG125	L2a	16086C 16223T 16278T 16294T 16390A 73G 146C 152C 195C 198T 263G 315.1C
FS_ANG126	T2c1	16126C 16292T 16294T 16296T 16519C 73G 146C 263G 315.1C 523DEL 524DEL
FS_ANG127	L2a1	16223T 16234T 16249C 16278T 16294T 16295T 16390A 16526A 73G 143A 146C 152C 195C 263G 315.1C
FS_ANG128	M1b2	16129A 16183C 16189C 16223T 16249C 16311C 16399G 16519C 73G 195C 263G 315.1C 489C 16093C 16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523DEL 524DEL
FS_ANG129	L3e3	16093C 16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523DEL 524DEL
FS_ANG130	L3e2b	16172C 16183C 16189C 16192T 16223T 16320T 16519C 73G 150T 152C 195C 263G 309.1C 315.1C
FS_ANG131	L3e2b	16172C 16183C 16189C 16223T 16320T 16519C 73G 150T 152C 195C 263G 309.1C 315.1C
FS_ANG132	L3e2b	16172C 16183C 16189C 16223T 16320T 16366T 16519C 73G 150T 195C 263G 309.1C 315.1C 573.1C 573.2C
FS_ANG133	H24	16293G 16362C 16519C 263G 309.1C 309.2C 315.1C
FS_ANG134	L2a1	16189C 16192T 16223T 16278T 16294T 16309G 16390A 16519C 73G 146C 152C 195C 263G 309.1C 315.1C
FS_ANG135	L2a1a1	16223T 16278T 16294T 16309G 16368C 16390A 16519C 73G 146C 152C 195C 263G 309.1C 309.2C 315.1C 524.1A 524.2C 524.3A 524.4C
FS_ANG136	L1c1d	16038C 16086C 16129A 16187T 16189C 16223T 16278T 16284G 16293G 16294T 16311C 16360T 16519C 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 309.1C 315.1C 316A 523DEL 524DEL
FS_ANG137	L2a1'2'3'4	16131C 16189C 16223T 16234T 16278T 16294T 16390A 73G 146C 152C 263G 309.1C 309.2C 315.1C
FS_ANG138	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG139	L2c2b1b	16069T 16223T 16234T 16264T 16278T 16311C 16390A 16519C 73G 93G 146C 150T 152C 182T 183G 195C 198T 263G 315.1C 325T 523DEL 524DEL
FS_ANG140	L3e1a	16182T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG141	L3f2/L3h	16192T 16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG142	L3d1a1a	16124C 16223T 16311C 16319A 73G 150T 152C 263G 315.1C 523DEL 524DEL
FS_ANG143	L3e3	16172C 16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523DEL 524DEL
FS_ANG144	L1c1a	16093C 16129A 16187T 16189C 16223T 16263C 16278T 16293G 16294T 16311C 16360T 16368C 16519C 73G 151T 152C 182T 186A 189C 195C 247A 263G 297G 315.1C 316A 523DEL 524DEL
FS_ANG145	L0a2a2	16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 204C 207A 236C 247A 263G 315.1C 523DEL 524DEL
FS_ANG146	L3f2	16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 315.1C 16124C 16183C 16189C 16223T 16234T 16278T 16304C 16311C 73G 152C 263G 309.1C 309.2C 315.1C 523DEL 524DEL
FS_ANG147	L3d3a	16124C 16183C 16189C 16223T 16234T 16278T 16304C 16311C 73G 152C 263G 309.1C 309.2C 315.1C 523DEL 524DEL
FS_ANG148	L1c3b'c	16129A 16187T 16183C 16189C 16223T 16278T 16293G 16294T 16311C 16360T 16519C 16527T 73G 151T 152C 182T 186A 189C 247A 263G 265C 315.1C 316A 523DEL 524DEL
FS_ANG149	L3e2b	16172C 16183C 16189C 16223T 16320T 16519C 73G 150T 195C 263G 315.1C
FS_ANG150	L0a2	16093C 16148T 16172C 16187T 16188G 16189C 16223T 16230G 16311C 16320T 16519C 64T 93G 152C 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG151	L3e1a	16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG152	L0a1b1	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG153	L2a1a3c	16093C 16223T 16256T 16278T 16294T 16309G 16390A 16519C 73G 143A 146C 152C 195C 263G 315.1C
FS_ANG154	L3e2b	16172C 16183C 16189C 16223T 16301T 16320T 16519C 73G 150T 195C 263G 309.1C 315.1C
FS_ANG155	L3d1a1a	16124C 16223T 16319A 73G 150T 152C 263G 315.1C 523DEL 524DEL
FS_ANG156	L1c3a1b	16129A 16183C 16189C 16215G 16223T 16278T 16294T 16311C 16355T 16360T 16519C 73G 151T 152C 182T 183G 186A 189C 204C 247A 257G 263G 309.1C 315.1C 316A 523DEL 524DEL
FS_ANG157	K2a1a	16066G 16182C 16183C 16189C 16223T 16224C 16278T 16390A 16519C 73G 146C 152C 182T 263G 309.1C 309.2C 315.1C 511T
FS_ANG158	L2c	16223T 16278T 16390A 73G 89C 93G 146C 150T 182T 195C 198T 263G 309.1C 315.1C 325T 523DEL 524DEL
FS_ANG159	L3e1a	16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 315.1C
FS_ANG160	L3e2b	16172C 16189C 16223T 16320T 16519C 73G 150T 152C 195C 263G 315.1C
FS_ANG161	L4b2	16223T 16233G 16271C 16293T 16311C 16355T 16362C 16399G 44.1C 73G 146C 152C 244G 263G 309.1C 315.1C
FS_ANG162	L1c	16093C 16129A 16183C 16189C 16215G 16223T 16278T 16293G 16294T 16311C 16327T 16355T 16360T 16390A 16519C 73G 151T 152C 182T 186A 189C 195C 247A 263G 309.1C 315.1C 316A
FS_ANG163	L1c2	16129A 16183M 16189C 16223T 16265C 16278T 16286G 16294T 16311C 16360T 16519C 16527T 73G 151T 152C 182T 186A 189C 195C 198T 247A 263G 297G 315.1C 316A
FS_ANG164	R0	16468C 16519C 263G 309.1C 315.1C
FS_ANG165	R0	16519C 185A 263G 309.1C 315.1C
FS_ANG166	L3e1a	16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG167	T2c1d	16093C 16126C 16292T 16294T 16519C 73G 146C 263G 279C 309.1C 315.1C
FS_ANG168	H11a	16293G 16311C 195C 263G 309.1C 315.1C
FS_ANG169	L2b1a	16114A 16129A 16213A 16223T 16278T 16355T 16362C 16390A 73G 150T 152C 182T 195C 198T 204C 263G 309.1C 315.1C 418T 523DEL 524DEL
FS_ANG170	L3a1a3	16093C 16223T 16278T 16294T 16390T 73G 143A 146C 152C 182T 195C 263G 309.1C 315.1C 523DEL 524DEL
FS_ANG171	L3e1a	16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C
FS_ANG172	L3e3	16223T 16265T 16519C 73G 150T 195C 263G 309.1C 315.1C 499A 523DEL 524DEL
FS_ANG173	T2b3	16126C 16294T 16296T 16304C 16519C 73G 151T 263G 315.1C
FS_ANG174	L3e1a1	16185T 16223T 16311C 16327T 73G 150T 185A 189G 263G 315.1C
FS_ANG175	L1c2	16129A 16187T 16189C 16223T 16265C 16286G 16294T 16311C 16360T 16519C 16527T 73G 151T 152C 182T 186A 189C 195C 198T 200G 247A 263G 297G 315.1C 316A
FS_ANG176	L2a1	16129A 16223T 16234T 16249C 16278T 16294T 16295T 16390A 73G 143A 146C 152C 195C 263G 309.1C 315.1C

Na tabela: A- Adenina, C- Citosina; G- Guanina; T- Timina; DEL- Deleção; M – Adenina e Citosina

4.1 Composição nucleotídica

Foram identificados 145 haplótipos, sendo 125 únicos, presentes uma única vez no conjunto de amostras estudadas. Os restantes 20 haplótipos encontram-se presente em mais do que uma amostra. Neste sentido, 14 sequências partilham o seu haplótipo em grupos de 2 amostras, 4 sequências em grupos de 3 amostras e 2 sequências em grupos de 4 amostras. Os dois haplótipos mais comuns, presentes em 4 sequências cada, estão descritos na Tabela 8.

Tabela 8 – Haplótipos de maior frequência nas 173 sequências estudadas no presente estudo

Amostras	Haplótipos
ANG078, ANG151, ANG166, ANG171	16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL
ANG012, ANG077, ANG106, ANG152	16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C

No que diz respeito às alterações nucleotídicas observadas relativamente à rCRS (Andrews *et al.* 1999), nas 173 sequências foram identificados 2 911 polimorfismos, distribuídos por 249 posições nucleotídicas, compreendidas entre as posições 16 024 e 576 do ADNmt (Tabela 9). De forma detalhada, 2 278 alterações correspondem a transições, sendo 1 472 entre pirimidinas e 806 entre purinas. Por outro lado, 174 alterações correspondem a transversões entre nucleótidos. Foram também identificadas 284 inserções e 174 deleções. Os polimorfismos descritos no grupo de sequências estudadas encontram-se discriminados na Tabela 10.

Tabela 9 – Distribuições dos polimorfismos observados na região controlo total do ADNmt nas 173 sequências de imigrantes Angolanos a residir em Lisboa.

	Nucleótido	Número de polimorfismo
Número de posições polimórficas		249
Transições entre pirimidinas	De C para T	766
	De T para C	706
Transições ente purinas	De A para G	234
	De G para A	572
Transversões	De C para A	40
	De T para A	4
	De A para C	89
	De T para G	1
	De C para G	26
	De A para T	7
	De G para T	7
Inserções	C	278
	A	6
Deleções	A	87
	C	87
Heteroplasmia de posição	De A para M	1

A rCRS apresenta uma região homopolimérica no segmento HVII composta por 5 Citosinas (C) entre as posições 310 e 315; em todas as sequências analisadas no presente estudo verificou-se a inserção de uma Citosina (C) na posição 315 que resulta na alteração **315.1C**. Uma outra mutação extremamente comum observada trata-se da transição de uma Adenina (A) por uma Guanina (G) na posição 263 (polimorfismo **263G**) e está ausente em apenas uma sequência – ANG073. As posições polimórficas descritas, observadas com elevada frequência no grupo de amostras estudado, não estão presentes na rCRS. No entanto, estes polimorfismos foram identificados como comuns por Anderson e colaboradores em 1981 (Anderson *et al.* 1981), e estão presentes em grande parte dos indivíduos.

A transição de uma Citosina (C) por uma Timina (T) na posição 16 223 da região HVI aparece em 153 sequências. Ainda, a transição de uma Adenina (A) por uma Guanina (G) na posição 73 da região HVII, está presente em 142 das 173 amostras estudadas. Na região HVII foi observada em 118 sequências a transição na posição 16 519 de uma Timina (T) por uma Citosina (C). Por outro lado, na região HVII, a Timina (T) da posição 152 é substituída por uma Citosina (C) em 103 sequências. Uma outra substituição comum é a transição de uma Citosina (C) por uma Timina (T) na posição 16 278 da região HVI, presente em 92 sequências.

O segmento localizado entre as posições 16 183 e 16 189 da região HVI, caracterizado pela sequência ACCCCCTCCCC, apresenta elevada variabilidade nucleotídica. Nesta região, foram identificadas 200 polimorfismos em 7 posições polimórficas. Na posição 16 183, 25 das sequências apresentam uma substituição da Adenina (A) por uma Citosina (C). A posição 16 185, que na sequência de referência é uma Citosina (C) é substituída em 7 sequências por uma Timina (T). Por outro lado, a posição 16 186 sofre apenas mutação numa das sequências estudadas, onde a Citosina (C) é substituída por uma Timina (T). Ainda, na posição 16 187, a Citosina (C) é substituída por uma Timina (T) em 50 sequências. A posição 16 188 sofre substituição da Citosina (C), novamente por uma Timina (T), em 22 sequências. Por outro lado a Timina (T) da posição 16 189 é substituída por uma Citosina (C) em 89 das 173 sequências estudadas. Por último na posição 16 192 da mesma região, a Citosina (C) é substituída por uma Timina (T) em 6 sequências.

A região do segmento HVII compreendida entre as posições 303 e 315 contém um conjunto de 7 Citosinas (C), uma Timina (T) e 5 Citosinas (C). Já foi referido anteriormente a elevada frequência da inserção de uma citosina (C) na posição 315. Nesta região é também comum a inserção de Citosinas (C) na posição 309. Em 67 das sequências esta inserção é de apenas um nucleótido, sendo identificada como **309.1C**. Em 10 sequências a inserção inclui duas Citosinas (C) e é classificada como **309.1C 309.2C**.

Entre as posições 515 e 523 da região HVIII estão presentes 5 repetições seguidas dos nucleótidos Adenina (A) e Citosina (C). Em 86 das 173 sequências estudadas (aproximadamente 50%) verifica-se a deleção de um dos conjuntos de Adenina (A) e Citosina

(C). De acordo com a nomenclatura proposta por Carracedo e colaboradores em 2000 (Carracedo *et al.* 2000), estas alterações são classificadas como **523Del 524Del**. Ainda associada a esta região, em 3 amostras, verificou-se a inserção de duas duplas de Adenina (A) e Citosina (C) na posição 524, que se classificou como **524.1A 524.2C 524.3A 524.4C**.

No final da região HVIII verificou-se a existência de inserções na posição 573 presente em 4 sequências. Em duas das amostras, verificou-se a inserção de 2 Citosinas (C) enquanto nas outras duas amostras verificou-se a inserção de 4 Citosinas (C). No primeiro caso a classificação desses polimorfismos é **573.1C 573.2C** e no segundo a classificação dos polimorfismos é **573.1C 573.2C 573.3C 573.4C**.

Nenhuma das sequências estudadas está em total concordância com a rCRS. O haplótipo que apresenta um menor número de polimorfismos inclui 3 alterações. Esta sequência está presente em duas amostras, sendo elas ANG033 e ANG094 e o respectivo haplótipo **16519C 263G 315.1C**. Contrariamente, a sequência que apresenta um maior número de polimorfismos é a ANG121, que contém um total de 30 alterações: **16129A 16182C 16183C 16189C 16223T 16278T 16293G 16294T 16311C 16355T 16360T 16519C 73G 89C 93G 95C 151T 152C 182T 186A 189C 195C 236C 247A 263G 297G 315.1C 316A 523DEL 524DEL**.

Tabela 10 – Polimorfismos presentes nas 173 sequências de ADNmt estudadas, compreendidos entre as posições 16 024 e 16 569.

		A	C	G	T			A	C	G	T			A	C	G	T
64	C				9	497	C				1	16213	G	5			
73	A			142		498	C		1d			16214	C				1
76	C				1	499	G	2				16215	A			5	
89	T		6			511	C				2	16222	C				1
93	A			33		513	G	1				16223	C				153
95	A		18			518	C				3	16224	T		3		
143	G	8				521	A			1		16225	C				2
146	T		43			523	A	86d				16230	A			23	
150	C				54	524	C		86d			16233	A			1	
151	C				32	524.1	-	3				16234	C				9
152	T		103			524.2	-		3			16241	A			4	
154	T		1			524.3	-	3				16243	T		1		
182	C				51	524.4	-		3			16249	T		4		
183	A			7		567	A			1		16256	C				7
185	G	17			6	573.1	-		4			16261	C				2
186	C	32				573.2	-		4			16263	T		3		
188	A			1		573.3	-		2			16264	C				12
189	A		32	42		573.4	-		2			16265	A		6		4
195	T		82			16038	A			3		16270	C				9
198	C				28	16051	A			1		16271	T		2		
199	T		1			16066	A			2		16274	G	7			
200	A			23		16069	C				2	16278	C				92
204	T		13			16071	C				1	16284	A			1	
207	G	10				16078	A			1		16286	C	1		4	2
228	G	1				16086	T		8			16287	C	1			
236	T		26			16092	T		2			16291	C				1
244	A			2		16093	T		11	1		16292	C				4
247	G	61				16114	C	6				16293	A			36	2
248	A			1		16124	T		14			16294	C				63
249	A	1d				16126	T		12			16295	C				3
257	A			2		16129	G	53				16296	C				3
263	A			172		16131	T		3			16301	C				2
264	C				2	16145	G	1				16304	T		5		
265	T		1			16148	C				22	16309	A			18	
272	A		1			16163	A			5		16311	T		86		
279	T		2			16164	A			1		16319	G	3			
295	C				1	16168	C				13	16320	C				37
297	A			20		16172	T		36			16327	C				16
309.1	-		77			16173	C				1	16354	C				1
309.2	-		10			16176	C				2	16355	C				7
315.1	-		173			16182	A		7		1	16356	T		1		
316	G	33				16183	A	1M	25			16360	C				32
325	C				6	16185	C				7	16362	T		9		
357	A			6		16186	C				1	16366	C				2
373	A			1		16187	C				50	16368	T		8		
385	A			1		16188	C			22		16390	G	37			1
416	T	4				16189	T		89			16399	A			6	
418	C				1	16192	C				6	16468	T		1		
458	C				1	16207	A			1		16519	T		118		
462	C				1	16209	T		10			16526	G	3			
489	T		2			16212	A			1		16527	C				7

Na tabela: d- deleção; A- Adenina; C- Citosina; G- Guanina; T- Timina; M- Citosina (C) e Adenina (A)

4.2 Heteroplasma

Nas 173 sequências de ADNmt analisadas, foram detetadas 55 amostras heteroplasmicas.

No que diz respeito à heteroplasma de comprimento, foram identificadas 66 heteroplasmias distribuídas por 55 amostras com heteroplasma de comprimento em pelo menos uma região hipervariável. Das 55 amostras, 25 apresentam heteroplasma de comprimento apenas no segmento HVI (entre as posições 16184 e 16193), 19 amostras apresentam heteroplasma de comprimento apenas no segmento HVII (entre as posições 303 e 309 e as posições 311 e 315) e uma amostra apresenta heteroplasma de comprimento apenas no segmento HVIII (entre as posições 568 e 573). Ainda, em 8 das amostras, verificou-se heteroplasma de comprimento na região HVI e na região HVII e em 2 amostras a heteroplasma de comprimento estava presente tanto na região HVI como na região HVIII. Cerca de 32% das sequências apresenta heteroplasma de comprimento, em pelo menos uma das três regiões hipervariáveis sequenciadas. A Tabela 11 resume a presença de heteroplasma de comprimento nas amostras estudadas.

Tabela 11 – Heteroplasma de comprimento observada nas 173 sequências de ADNmt de imigrantes Angolanos a residir em Lisboa.

	Segmento hipervariável	Frequência	%
Amostras homoplasmicas		118	68,21
Amostras heteroplasmicas	HVI	25	14,45
	HVII	19	10,98
	HVIII	1	0,58
	HVI e HVII	8	4,62
	HVI e HVIII	2	1,16
	Total	55	31,79

Para que sejam obtidos resultados com elevada confiança é necessária a aquisição de sequências de ADNmt amplificadas com o *primer* reverso e o *primer* direto. Nos casos em que se verificou heteroplasma de comprimento, não foi possível analisar a sequência a partir da posição em que o fenómeno de heteroplasma acontece pela sobreposição de picos no electroferograma adquirido, resultante de misturas de bases nucleotídicas em cada posição da sequência. Assim, foi necessário realizar novas fases de amplificação e sequenciação com o *primer* direto e o *primer* reverso. Este processo laboratorial permite obter duas cópias de cada sequência (até à posição heteroplasmica), o que garante a confirmação dos polimorfismos identificados.

Nas amostras ANG035, ANG104 e ANG105, o fenómeno de heteroplasma de comprimento está presente na região polimórfica de HVIII entre as posições 568 e 573. Sendo

este segmento localizado no final da região controlo total, a sequenciação adquirida com o *primer* reverso H639 não permite identificar nenhuma posição nucleotídica pela sobreposição de picos de bases diferentes observada desde o início da sequência. Nestes casos, a sequenciação das regiões HVII e HVIII foi realizada com duplicação da amplificação com *primer* direto L16555.

Verificou-se heteroplasmia de posição na posição 16 183 da amostra ANG163, onde é visível uma adenina (A) e uma citosina (C). De acordo com o código estabelecido pela IUPAC, a heteroplasmia é classificada como **16183M**.

4.3 Haplogrupos

Os 145 haplótipos identificados foram inseridos em 81 haplogrupos distintos. Esta classificação tem por base os polimorfismos presentes em cada segmento da região controlo total do ADNmt relativamente à rCRS e é realizada a partir *PhyloTree* (van Oven & Kayser 2009). Cada haplogrupo é identificado com letras do alfabeto seguida de números, que permite uma subdivisão de subhaplogrupos. Para que uma sequência seja incluída num ramo da árvore filogenética, precisa conter os polimorfismos presentes nesse mesmo ramo; contudo pode apresentar outros polimorfismos para além daqueles que permitem a sua inclusão num determinado haplogrupo. Neste sentido, duas sequências semelhantes que apresentam apenas alguns polimorfismos distintos, podem ser incluídas num mesmo haplogrupo. Como exemplo têm-se as amostras ANG090 e ANG129 cujos haplótipos são, respetivamente, **16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523DEL 524DEL** e **16093C 16223T 16265T 16519C 73G 150T 195C 263G 315.1C 523DEL 524DEL**. Como se pode verificar, a diferença entre ambas está na presença do polimorfismo **16093C** na amostra ANG129. Apesar disso, ambas se incluem no haplogrupos L3e3, por apresentarem as mutações **195C!** e **16265T**, necessárias à inclusão nesse ramo.

O haplótipo da amostra ANG141, apresenta polimorfismos que o permitem incluir em dois haplogrupos distintos - L3f2 e L3h1. O polimorfismo responsável pela inclusão do haplótipo em cada um dos haplogrupos é a alteração **16311C!**. Contudo, o haplótipo associado a esta amostra - **16192T 16209C 16223T 16311C 16519C 73G 150T 189G 200G 263G 309.1C 315.1C** -, para além da posição **16311C!** apresenta também a posição **16209C**, necessária à inclusão da sequência no haplogrupo L3f2. Uma vez que existe uma maior discriminação associada ao haplogrupo L3f2 considera-se este, como sendo aquele em que o haplótipo da amostra ANG141 se inclui.

Do grupo de amostras estudadas, 151 sequências (cerca de 87%), pertencem ao macrohaplogrupo L. Este macrohaplogrupo, assim como as suas ramificações, são característicos de populações Africanas. Menos representativos são os restantes haplogrupos descritos na amostragem estudada - H, J, K, R0, M, T e U - característicos da região

euroasiática e que representa cerca de 13% das sequências estudadas. A distribuição dos diferentes haplogrupos encontra-se descrita na Figura 21.

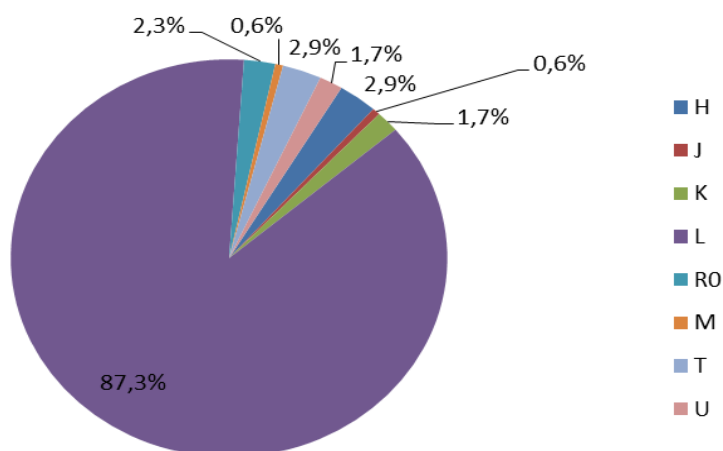


Figura 21 – Percentagem dos diferentes haplogrupos identificados nas 173 sequências de ADNmt pertencentes a imigrantes Angolanos a residir em Lisboa.

4.4 Cálculos da diversidade

A diversidade de sequência (\hat{H}) (*gene diversity*) é definida como a probabilidade de dois haplótipos escolhidos aleatoriamente serem diferentes. Este valor é calculado a partir da seguinte fórmula (Nei & Tajima 1981) :

$$(A): \hat{H} = \frac{n}{n-1} \left(1 - \sum_{i=1}^k p_i^2 \right)$$

Na fórmula, **n** representa o número de amostras em estudo, **k** o número de haplótipos e p_i^2 a frequência dos diferentes haplótipos presentes na população estudada. O valor da frequência os haplótipos é o número de vezes que uma determinada sequência se repete no conjunto de amostras. Desta forma, o valor de n será 173 e k será 145. Para determinar o somatório dos diferentes valores de p_i^2 foi necessário realizar o cálculo para cada frequência de haplótipos. Este valor é obtido pelo número de vezes que a sequência ocorre, a dividir pelo número total de indivíduos estudados, por sua vez a multiplicar pelos haplótipos que se repetem um determinado número de vezes. Estes valores encontram-se calculados e representados na Tabela 12.

Tabela 12 – Cálculo do valor de π^2 para a totalidade da região controle do ADNmt.

Nº de repetições	Cálculo	π^2
1	$(1/173)*125$	0,004177
2	$(2/173)*14$	0,001871
3	$(3/173)*4$	0,001203
4	$(4/173)*2$	0,001069
Somatório		0,00832

Pela aplicação da fórmula, o valor de \hat{H} será obtido da seguinte forma:

$$\hat{H} = 173 / (173-1) * (1-0,00832) = \mathbf{0,997446}$$

Este valor revela ser inferior quando calculado a partir do programa Arlequin, sendo $\hat{H}=0,9964$.

O cálculo foi efetuado para a mesma população, sendo consideradas apenas as regiões HVI e HVII e o segmento HVI isolado. O valor obtido a partir do *software* Arlequin foi $\hat{H}=0,9959$ e $\hat{H}=0,9933$, respetivamente.

O índice de diversidade nucleotídica (*nucleotide diversity*) é definido como número de diferenças, por nucleótido, entre a sequência em estudo e a sequência de referência. Este valor é representado por π e é calculado de acordo com a formula proposta por Nei e Tajima em 1981 (Nei & Tajima 1981):

$$(B): \quad \pi = (n/n-1) * (1/l) * \sum (1-x_i^2)$$

O n representa o total de indivíduos estudados (173 sequências), l o tamanho da sequência estudada (1122 pb) e x_i^2 a frequência relativa de cada nucleótido na posição i . O cálculo de x_i^2 é realizado a partir do valor da frequência relativa de cada nucleótido. Este por sua vez é calculado com a frequência absoluta sobre o número de amostras. A frequência absoluta de cada nucleótido será o número de vezes que esse ocorre como mutação na amostragem (substituição, deleção e inserção) a dividir pelo número de haplótipos identificados (145 haplótipos). Os valores determinados para cada nucleótido encontram-se descritos na Tabela 13. Para a realização dos cálculos não foi considerada a posição heteroplasmica **16183M** da amostra ANG163.

Tabela 13 – Cálculo do valor de x_i^2 para a totalidade da região controle do ADNmt.

Nucleótidos	A	C	G	T
Frequência absoluta	2,613793103	8	4,075862	5,37931
Frequência relativa	0,015108631	0,046243	0,02356	0,031094
x_i^2	0,000228271	0,002138	0,000555	0,000967

O valor de somatório $\Sigma (1-x_i^2)$ será:

$$(1-0,000228) + (1-0,002138) + (1-0,00555) + (1-0,000967) = 3.996111$$

Por aplicação da fórmula, o valor da diversidade nucleotídica será $\pi=0.003582$.

Este cálculo foi também realizado através do programa Arlequin sendo o valor obtido para diversidade nucleotídica $\pi=0.016028$.

O parâmetro de diversidade foi também obtido para os segmentos HVI e HVII e para o segmento HVI apenas, sendo os resultados calculados $\pi=0,025058$ e $\pi=0,017601$, respetivamente.

Estão resumidos nas tabelas 14, 15 e 16 os valores diversidade nucleotídica e diversidade de sequência da população estudada e de outras descritas na literatura, correspondentes à região controlo total do ADNmt (Tabela 14), aos segmentos HVI e HVII (Tabela 15) e ao segmento HVI (Tabela 16). Os valores de parâmetros de diversidade da população imigrante Angolana a residir em Lisboa apresentados nas tabelas são aqueles obtidos a partir do *software* Arlequin, uma vez que os calculados para as outras populações também foram efetuados através do mesmo programa. Ocorre a exceção com os valores de Cabo Verde na Tabela 14 onde são utilizados os resultados obtidos manualmente pelo autor (Morais 2013).

Tabela 14 – Valores de diversidade de sequência (\hat{H}) e diversidade nucleotídica (π) obtidos para a totalidade da região controlo do ADNmt.

População	N	\hat{H}	π	Referência
Imigrantes Angolanos	173	0,996	0.016	---
Babinga (República do Congo)	23	0.889	0.009	(Batini <i>et al.</i> 2011)
Baka (Camarão)	27	0.855	0.013	(Batini <i>et al.</i> 2011)
Baka (Gabão)	20	0.889	0.007	(Batini <i>et al.</i> 2011)
Bakola (Camarão)	11	0.909	0.007	(Batini <i>et al.</i> 2011)
Biaka (República Centro Africana)	23	0.968	0.011	(Batini <i>et al.</i> 2011)
Mbenzele (República Centro Africana)	30	0.954	0.007	(Batini <i>et al.</i> 2011)
Mbuti (República Democrática do Congo)	14	0.934	0.011	(Batini <i>et al.</i> 2011)
Mbuti (República Democrática do Congo)	21	0.981	0.011	(Batini <i>et al.</i> 2011)
Fang (Gabão)	17	1.000	0.016	(Batini <i>et al.</i> 2011)
Nzebi (Gabão)	19	0.977	0.016	(Batini <i>et al.</i> 2011)
Cabo Verde	103	0,992	0,003	(Morais 2013)

Tabela 15 - Valores de diversidade de sequência (\hat{H}) e diversidade nucleotídica (π) obtidos para os segmentos HVI e HVII da região controle do ADNmt.

População	N	\hat{H}	π	Referência
Imigrantes Angolanos	173	0,996	0,025	--
Bantu Angola	44	0,997	---	(Plaza <i>et al.</i> 2004)
Mozambique	109	0,973	---	(Pereira <i>et al.</i> 2001)

Tabela 16 - Valores de diversidade de sequência (\hat{H}) e diversidade nucleotídica (π) obtidos para o segmento HVI da região controle do ADNmt.

População	N	\hat{H}	π	Referência
Imigrantes Angolanos	173	0,993	0,018	--
Bantu (Bakongo)	70	1,000	0,016	(Carvalho <i>et al.</i> 2011)
Bantu (Kibundu)			0,017	
Bantu (Ovimbundo)			0,016	
Bantu (Kuvale)	54	0,937	0,026	(Coelho <i>et al.</i> 2009)
Bantu (Ganguela)	21	0,962	0,025	(Coelho <i>et al.</i> 2009)
Bantu (Nhaneca-Humbe)	153	0,982	0,024	(Coelho <i>et al.</i> 2009)
Bantu (Ovimbundo)	92	0,987	0,024	(Coelho <i>et al.</i> 2009)
Bantu (Shona)	59	0,978	---	(Castrì <i>et al.</i> 2009)
Bantu (Hutu)	42	0,979	---	(Castrì <i>et al.</i> 2009)
São Tomé e Príncipe	103	0,912	---	(Trovoada <i>et al.</i> 2004)
Bantu Angola	44	0,992	---	(Plaza <i>et al.</i> 2004)
Moçambique	109	0,962	---	(Pereira <i>et al.</i> 2001)
Gana	191	0,987	---	(Fendt, Röck, <i>et al.</i> 2012)
Guiné Bissau	88	0,991	---	(Carvalho <i>et al.</i> 2011)
Cabo Verde	103	0,984	0,011	(Morais 2013)

4.5 Análise filogenética

Através dos haplótipos obtidos foi possível realizar comparação da população em estudo, com resultados já publicados para outras populações Africanas. Foi aplicada a AMOVA, que permite calcular a variabilidade genética entre populações. Para tal foram selecionadas da literatura, populações previamente estudadas, descritas na Tabela 17.

Tabela 17 – Populações retiradas da literatura referentes à totalidade da região controlo do ADNmt, utilizadas para fins comparativos.

Populações	Tamanho da amostra	Referências
Imigrantes Angolanos	173	---
Angola (Khoe-san)	129	(Fendt, Huber, <i>et al.</i> 2012)
Gana	191	(Fendt, Röck, <i>et al.</i> 2012)
Somália	190	(Mikkelsen <i>et al.</i> 2012)
Cabo Verde	103	(Morais 2013)

Foram calculados os valores de F_{st} e de **p-value** entre grupos de duas populações (Tabela 18), tendo sido posteriormente efetuada a representação gráfica com base no cálculo das distâncias entre populações (Figura 22).

Tabela 18 – Valores de F_{st} e p obtidos através do programa Arlequin para a totalidade da região controlo do ADNmt.

Populações	Imigrantes Angolanos	Cabo Verde	Angola (Khoe-san)	Gana	Somália
Imigrantes Angolanos	---	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000
Cabo Verde	0.04408	---	0.00000+-0.0000	0.00000+-0.0000	0.00000+-0.0000
Angola (Khoe-san)	0.24677	0.30102	---	0.00000+-0.0000	0.00000+-0.0000
Gana	0.02746	0.02139	0.27996	---	0.00000+-0.0000
Somália	0.03588	0.06419	0.28884	0.05876	---

As sequências foram reduzidas aos segmentos HVI e HVII, tendo sido a AMOVA aplicada de igual forma, a um conjunto de populações descritas na literatura, apenas para as regiões HVI e HVII sequenciadas (Tabela 19). Os parâmetros de F_{st} e p-value calculados e a representação gráfica das distâncias calculadas encontram-se na Tabela 20 e Figura 23.

Tabela 19 - Populações retiradas da literatura referentes aos segmentos HVI e HVII da região controlo do ADNmt, utilizadas para fins comparativos.

Populações	Tamanho da amostra	Referências
Imigrantes Angolanos	173	---
Angola (Khoe-san)	129	(Fendt, Huber, <i>et al.</i> 2012)
Gana	192	(Fendt, Röck, <i>et al.</i> 2012)
Somália	190	(Mikkelsen <i>et al.</i> 2012)
Cabo Verde	103	(Morais 2013)
Guiné Bissau	88	(Carvalho <i>et al.</i> 2011)
Moçambique	109	(Pereira <i>et al.</i> 2001)
São Tomé e Príncipe	100	(Trovoada <i>et al.</i> 2004)
Angola	59	(Carvalho <i>et al.</i> 2006)
Bantu (outros)	45	(Coelho <i>et al.</i> 2009)
Bantu (Kuvale)	54	(Coelho <i>et al.</i> 2009)
Bantu (Guanguela)	21	(Coelho <i>et al.</i> 2009)
Bantu (Nhaneca-Humbe)	153	(Coelho <i>et al.</i> 2009)
Bantu (Ovimbundu)	92	(Coelho <i>et al.</i> 2009)
Cabinda	110	(Beleza <i>et al.</i> 2005)

Tabela 20 - Valores de Fst e p obtidos através do programa Arlequin para os segmentos HVI e HVII da região controlo do ADNmt.

Populações	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	---	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.12613+- 0.0278	0.39640+- 0.0338	0.0000+- 0.0000	0.40541+- 0.0365	0.0000+- 0.0000	0.19820+- 0.0402	0.16216+- 0.0379	0.48649+- 0.0594	0.03604+- 0.0148
2	0.01868	---	0.40541+- 0.0529	0.0000+- 0.0000	0.03604+- 0.0201	0.0000+- 0.0000	0.01802+- 0.0121	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.00901+- 0.0091	0.0000+- 0.0000	0.00901+- 0.0091	0.0000+- 0.0000
3	0.04056	0.00028	---	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000
4	0.03909	0.04571	0.05973	---	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000
5	0.03065	0.01037	0.01712	0.06392	---	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000
6	0.03860	0.04267	0.06550	0.06691	0.03772	---	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.00901+- 0.0091	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000
7	0.00459	0.01483	0.02964	0.05934	0.02472	0.05353	---	0.40541+- 0.0493	0.0000+- 0.0000	0.21622+- 0.0243	0.0000+- 0.0000	0.03604+- 0.0148	0.06306+- 0.0237	0.45045+- 0.0271	0.03604+- 0.0148
8	0.0000	0.01840	0.03953	0.04612	0.02274	0.03269	0.00071	---	0.0000+- 0.0000	0.84685+- 0.0279	0.0000+- 0.0000	0.19820+- 0.0227	0.28829+- 0.0504	0.63964+- 0.0345	0.11712+- 0.0273
9	0.24421	0.27967	0.31074	0.28938	0.28299	0.26332	0.25542	0.26087	---	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000
10	0.0000	0.02003	0.05302	0.06079	0.02764	0.03440	0.00195	0.00000	0.25283	---	0.0000+- 0.0000	0.32432+- 0.0388	0.25225+- 0.0326	0.65766+- 0.0508	0.18018+- 0.0332
11	0.07913	0.13755	0.18906	0.16290	0.17124	0.13961	0.09081	0.08102	0.20030	0.05899	---	0.08108+- 0.0286	0.0000+- 0.0000	0.0000+- 0.0000	0.0000+- 0.0000
12	0.01005	0.06160	0.10226	0.07988	0.07928	0.06018	0.02629	0.01014	0.25746	0.00188	0.02419	---	0.19820+- 0.0485	0.30631+- 0.0492	0.34234+- 0.0504
13	0.00323	0.02854	0.05189	0.05117	0.04485	0.04637	0.00638	0.00325	0.21993	0.00288	0.05957	0.00762	---	0.42342+- 0.0560	0.11712+- 0.0273
14	0.0000	0.01823	0.03791	0.05043	0.02863	0.03993	0.0000	0.0000	0.24105	0.0000	0.07573	0.00492	0.00033	---	0.32432+- 0.0473
15	0.00744	0.04394	0.06650	0.06571	0.05865	0.06818	0.00791	0.00451	0.24655	0.00499	0.06015	0.00400	0.00334	0.00120	---

Na tabela: 1 – Imigrantes Angolanos; 2 – Guiné Bissau; 3 – Cabo Verde; 4 – Somália; 5 – Gana; 6 – Moçambique; 7 – São Tomé e Príncipe; 8 – Angola; 9 – Angola (Khoe-san); 10 – Bantu (outros); 11 – Bantu (Kuvale); 12 – Bantu (Ganguela); 13 – Bantu (Nhaneca-Humbe); 14 – Bantu (Ovimbundo); 15 - Cabinda

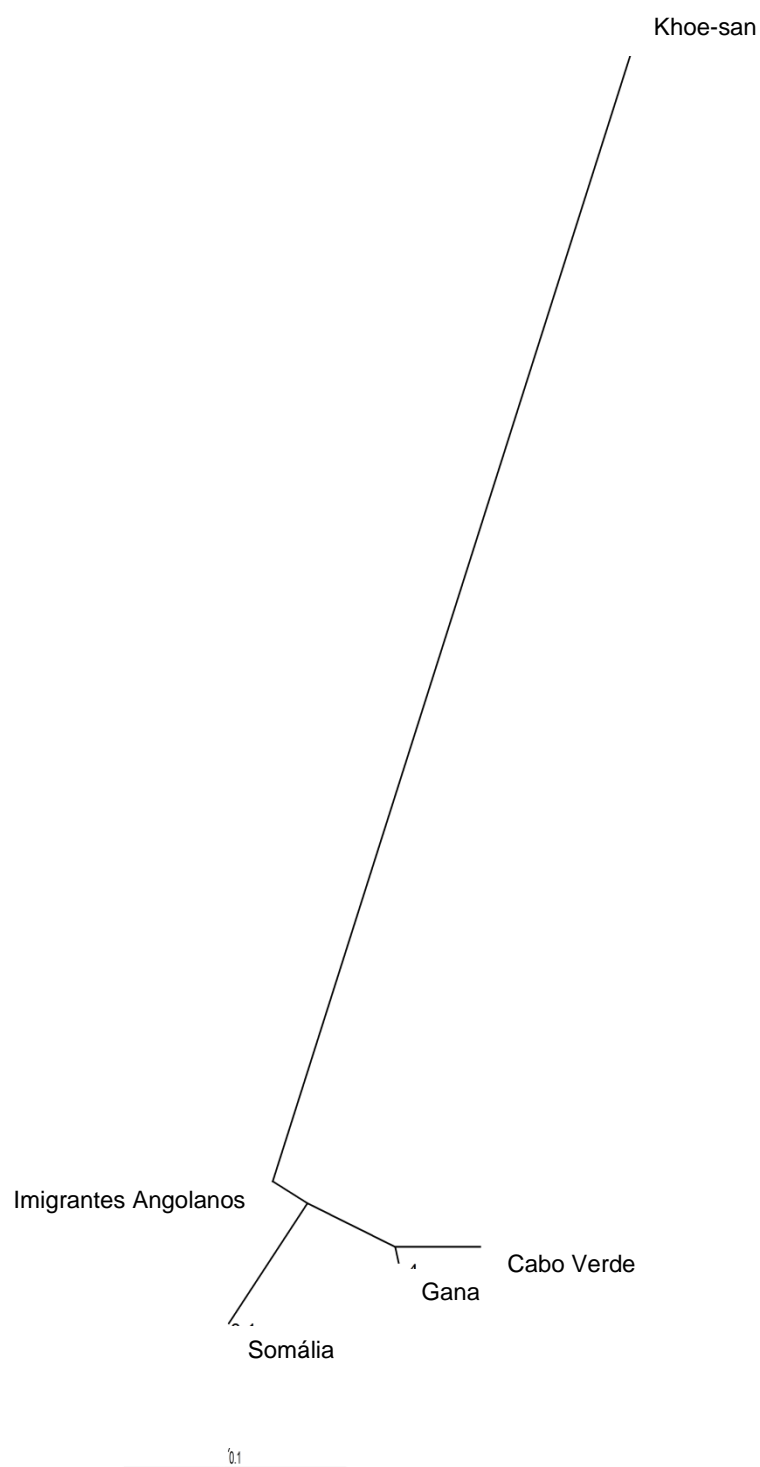


Figura 22 – Representação filogenética de distâncias genéticas de populações Africanas relativas à região controlo do ADNmt.

5. Discussão

A teoria da origem do Homem moderno em África é amplamente aceite na comunidade científica facto este que se tem traduzido num crescente interesse no estudo de diversas populações Africanas. Estas demonstram uma elevada heterogeneidade genética quando comparadas com populações de outras origens e apresentam uma história complexa, que os estudos genéticos podem ajudar a esclarecer. Investigações que têm por base o estudo de marcadores genéticos não recombinantes têm-se revelado importantes, uma vez que permitem completar informação arqueológica e linguística sobre o continente Africano. Desta forma, tanto o ADNmt como a zona não recombinante do cromossoma Y têm vindo a ser utilizados na caracterização de populações Africanas.

A ausência de recombinação, o elevado número de cópias, a linhagem materna e a elevada taxa de mutação, são características que fazem do ADNmt um marcador genético comumente utilizado em estudos de evolução e história humana. No presente estudo foi analisada a região controlo total do ADNmt, que inclui os três segmentos hipervariáveis – HVI, HVII e HVIII – que apresentam elevada taxa de mutação, comparativamente ao restante genoma.

Até à data não se encontram disponíveis quaisquer publicações referentes à região controlo total do ADNmt da população Angolana, a residir em Angola. Contudo existem publicações referentes aos segmentos HVI e HVII da região controlo do ADNmt de indivíduos Angolanos a residir em Angola e de diferentes grupos étnico-linguísticos Bantu. Plaza e colaboradores (2004) analisaram os segmentos HVI e HVII de 44 sequências de indivíduos pertencentes a dois grupos étnico-linguístico do povo Bantu residentes em Angola. Em 2005, Beleza e colaboradores publicaram informação, também referente aos segmentos HVI e HVII, da população de Cabinda, que apesar de pertencer a Angola, se encontra delimitada por território da República Democrática do Congo; um total de 109 sequências foi examinado por estes autores. No ano de 2006, Carvalho e colaboradores publicaram sequências das regiões HVI e HVII de ADNmt de 59 indivíduos Angolanos residentes em Angola. Mais tarde, em 2009, Coelho e seus colaboradores estudaram um total de 369 indivíduos pertencentes a diversos grupos étnico-linguístico do povo Bantu em Angola; as análises incluíram o estudo dos segmentos HVI e HVII do ADNmt assim como da região não recombinante do cromossoma Y.

5.1 Região controlo total do ADNmt

O estudo da totalidade da região controlo do ADNmt, compreendida entre as posições 16 024 e 576, num comprimento total de 1122pb, apresenta vantagens comparativamente às técnicas anteriormente utilizadas, por permitir uma classificação mais clara e detalhada dos haplogrupos nos quais os diversos haplótipos se inserem. Neste estudo foram utilizados dois pares de *primers* – L15997/H016 e L16555/H639 - que garantiram a amplificação da região controlo total do ADNmt. A amplificação desta região poderia ser realizada com apenas um par de *primers* que compreendesse um único segmento mais amplo. Contudo, a técnica utilizada no presente estudo revela ser de igual forma eficiente, mais económica e é realizada num curto período de tempo.

Os haplótipos analisados e determinados com apenas o estudo do segmento HVI do ADNmt, podem sofrer alterações quando se inclui informação referente à região HVII. Mutações compreendidas entre as posições 73 e 340 (segmento HVII) podem garantir a clarificação e distinção entre diversos haplótipos. A limitação do estudo ao segmento HVI traduz-se num número superior de haplótipos partilhados, revelando a sua menor eficiência.

Para exemplificar a baixa fidelidade do estudo do segmento HVI da região controlo do ADNmt, são apresentadas as sequências ANG131 e ANG132, do presente estudo. Os seus haplótipos **16172C 16183C 16189C 16223T 16320T 16519C 73G 150T 152C 195C 263G 309.1C 315.1C** e **16172C 16183C 16189C 16223T 16320T 16366T 16519C 73G 150T 195C 263G 309.1C 315.1C 573.1C 573.2C**, respetivamente. Com a análise apenas do segmento HVI apresentariam ambas o haplótipo **16172C 16183C 16189C 16223T 16320T**, com polimorfismos idênticos.

A amostra ANG011 que pertence ao haplogrupo H é caracterizada pelo haplótipo **16519C 152C 263G 309.1C 315.1C**. Nenhum dos polimorfismos apresentados pertence à região HVI, pelo que, se o estudo fosse realizado apenas com este segmento, o haplótipo desta sequência seria idêntico à rCRS, o que na realidade não acontece.

Por outro lado, com apenas o estudo das regiões HVI e HVII, é excluída ainda a análise do segmento HVIII da região controlo do ADNmt e os segmentos intermédios de HVI, HVII e HVIII, que apesar de não apresentarem uma taxa de mutação tão elevada como as regiões hipervariáveis, permitem também a discriminação dos haplótipos.

Na totalidade da região controlo do ADNmt, excetuando as regiões HVI e HVII, foram identificados 417 polimorfismos, distribuídos por 35 locais polimórficos, que representam cerca de 14% das posições que contêm polimorfismos nas sequências estudadas. Neste grupo de mutações estão incluídos 196 polimorfismos presentes nos segmentos da região controlo total do ADNmt que não se incluem nas regiões hipervariáveis.

A utilização da região controlo total do ADNmt permite distinguir linhagens maternas, que de outra forma, apenas com o estudo das regiões HVI e HVII, seriam classificadas como iguais.

As amostras ANG018 e ANG104 são um exemplo da vantagem de utilização da região controlo total nos estudos de ADNmt. Sem a análise da região controlo total, não seriam identificados os polimorfismos **523Del** e **524Del** presentes na amostra ANG018, assim como as alterações **5731.C**, **573.2C**, **573.3C** e **573.4C** presentes na amostra ANG104. Sendo que as restantes posições polimórficas de ambas as sequências, são coincidentes, estas duas amostras iriam apresentar o mesmo haplótipo. Com o estudo da região controlo total, foi possível distinguir duas linhagens maternas distintas, cada uma associada a uma amostra.

A análise da totalidade da região controlo do ADNmt garante uma maior discriminação e eficiência na atribuição dos haplogrupos às sequências determinadas. Para inclusão de haplótipos em diversos ramos da árvore filogenética de ADNmt é necessária a identificação de polimorfismos que apenas com a totalidade da região controlo são identificados.

Qualquer sequência que, através da análise da região controlo total, tenha sido atribuída ao haplogrupo L2 ou suas ramificações, tem que ser classificada no ramo L2'3'4'6, quando apenas os segmentos HVI e HVII são considerados. Isto acontece uma vez que para a sua inclusão no haplogrupo L2 são necessárias as mutações **146C!** **150T** **152C!** **16311T** **16390A**. A posição 16390 não pertence aos segmentos HVI e HVII. Contudo, muitos autores, quando restringem o estudo das suas linhagens maternas a estas duas regiões, garantem a sequenciação das amostras até à posição 16 390.

A amostra ANG021, com a análise da região controlo total, inclui-se no haplogrupo L1b1a10. Contudo, se o estudo fosse efetuado apenas com as regiões HVI e HVII, a sua classificação seria L1'2'3'4'5'6, uma vez que para a sua inclusão no haplogrupo L1b1 é necessária informação referente à posição 357, que sofre uma transição de Adenina (A) para Guanina (G). Sendo que a região HVII termina na posição 340, não seria possível obter informação relativa a este polimorfismo.

Um outro exemplo que demonstra a importância do estudo da região controlo total do ADNmt é aqui representado pela amostra ANG031 que se inclui no haplogrupo T2c1. Haveria a possibilidade de inserção da sequência no haplogrupo T2c1a, se esta apresentasse uma ou mais inserções de Citosinas (C) na posição 573. Sem o estudo da região HVIII, não seria possível esclarecer esta dúvida. Neste caso, com a análise da região HVIII realizada, verificou-se que esta sequência não apresenta qualquer inserção na posição 573, ficando a classificação do haplogrupo confinada a T2c1.

Relativamente à amostra ANG039, a sua classificação no haplogrupo L0a2 não seria possível uma vez que no estudo das regiões HVI e HVII apenas se consideram os intervalos de posições 16024-16365 e 73-340, excluindo desta forma a posição 64, onde a existência de um polimorfismo é essencial para a sua inclusão no haplogrupo referido.

Para que a amostra ANG042 seja inserida no haplogrupo L1c3c, é necessária a presença da mutação **458T**, pertencente à região HVIII. Apenas com o estudo de HVI e HVII, esta sequência seria classificada como pertencendo ao haplogrupo L1c3b'c.

A linhagem ANG043, cujo haplogrupo é J1c2c, precisa da posição **462T** para a sua distinção para o haplogrupo J1, o que apenas é possível averiguar com a região controlo total. Também, para que esta se inserisse no haplogrupo J, seria necessária a transição para uma Citosina (C) na posição 489. Neste sentido, se a análise se restringisse aos segmentos HVI e HVII, esta amostra seria inserida no haplogrupo R.

As sequências ANG085, ANG088, ANG118, ANG163 e ANG175, pertencem ao haplogrupo L1c2, para o qual é necessária a mutação **16527T**. Sem o estudo da região controlo total estas linhagens ficariam confinadas ao haplogrupo L1c1'2'4'6.

A *PhyloTree* para além dos polimorfismos associados à região controlo total do ADNmt, considera ainda aqueles presentes na região codificante, entre as posições 577 e 16023. As alterações presentes neste segmento contribuem de igual forma para a discriminação de determinados haplogrupos. Como exemplo tem-se a amostra ANG044 que pertence ao haplogrupo L3b. Para que esta seja inserida neste haplogrupo terá que incluir as seguintes mutações **3450T 5773A 6221C 9449T 10086G 13105G! 13914A 15311G 15824G 16124C 16278T! 16362C**. Apenas as últimas três (**16124C 16278T! 16362C**) fazem parte da região controlo, sendo as restantes pertencentes à região codificante do ADNmt. Assim, com o estudo do comprimento total da molécula de ADNmt, seria possível agrupar o haplótipo referente à amostra ANG044 no haplogrupo L3b, com uma maior confiança. Por outro lado, o estudo dos polimorfismos pertencentes à região codificante do ADNmt permitiria a inclusão de haplótipos em possíveis subhaplogrupos, garantindo, assim, um maior poder de discriminação.

Através dos exemplos demonstrados confirma-se a utilidade do estudo da região controlo total do ADNmt, que permite a descrição mais pormenorizada dos haplótipos e, consequentemente, uma maior discriminação dos haplogrupos e subhaplogrupos a que as sequências pertencem. O estudo da região codificante do genoma mitocondrial iria contribuir com informação adicional aos haplótipos determinados, permitindo uma maior qualidade e rigor nos resultados apresentados.

5.2 Discussão dos resultados obtidos

A introdução dos haplótipos na base de dados EMPOP revela ser de extrema importância uma vez que garante a disponibilidade das sequências de ADNmt para toda a comunidade científica, contribuindo com a ampliação da informação da base de dados de ADNmt.

5.2.1 Composição nucleotídica

De acordo com os resultados obtidos verificou-se um número superior de substituições do que de deleções e inserções estando de acordo com o descrito anteriormente (Carvalho *et al.* 2003). Segundo Budowle e colaboradores (1999) (Budowle *et al.* 1999) o número médio de polimorfismos presentes em populações de descendência materna Africana é de 15; o presente estudo exhibe uma média de 17 mutações por amostra, estado próximo do descrito por estes autores.

Algumas das mutações identificadas, apresentam baixa frequência no conjunto de 173 sequências analisadas. Por exemplo, a transição de uma Citosina (C) por uma Timina (T) na posição 76 está presente uma única vez. Este fenómeno é visível noutras posições polimórficas, descritas na Tabela 21.

Tabela 21 – Mutações que apresentam uma baixa frequência na população de imigrantes Angolanos a residir em Lisboa.

		A	C	G	T			A	C	G	T
76	C				1	16051	A			1	
154	T		1			16071	C				1
188	A			1		16078	A			1	
199	T		1			16145	G	1			
228	G	1				16164	A			1	
248	A			1		16173	C				1
249	A	1				16186	C				1
265	T		1			16468	T		1		
272	A		1			16207	A			1	
295	C				1	16212	A			1	
373	A			1		16214	C				1
385	A			1		16222	C				1
418	C				1	16233	A			1	
458	C				1	16243	T		1		
462	C				1	16284	A			1	
497	C				1	16287	C	1			
498	C		1d			16291	C				1
513	G	1				16354	C				1
521	A			1		16356	T		1		
567	A			1							

Na tabela: d- deleção; A- Adenina; C- Citosina; G- Guanina; T- Timina

A base de dados EMPOP disponibiliza uma ferramenta através do EMPOP *DrawNetWork* v1.24., que permite a construção de *Quasi-median networks* (Zimmermann *et al.* 2012). Estes são usados para avaliar a qualidade dos dados de ADNmt obtidos. É possível examinar os haplótipos de ADNmt através da representação gráfica da estrutura genética das linhagens maternas. Este programa garante a deteção de idiosincrasias e possíveis artefactos que possam advir de problemas de sequenciação ou da interpretação de resultados. Aquando

a introdução dos dados para criação da representação gráfica, é possível selecionar diferentes filtros, os quais irão garantir a remoção de mutações comuns que não serão consideradas na construção do *Network*. O filtro EMPOPall_R11 inclui todas as mutações já observadas na base de dados EMPOP permitindo uma análise rápida dos dados, destacando apenas as mutações que ainda não foram observadas. Através deste filtro foi possível averiguar que todas das mutações observadas no presente estudo, mesmo as que apresentam baixa frequência (Tabela 21), foram identificadas em populações presentes na base de dados. A representação obtida agrupa todos os haplótipos num mesmo *cluster* (Figura 24), não havendo mutações desconhecidas no grupo de amostras estudadas, que originem ramificações no *Network*.

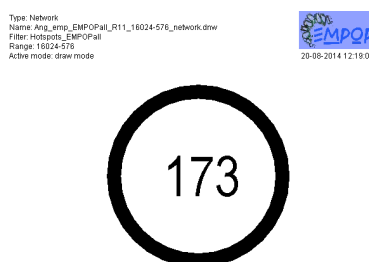


Figura 24 – *Network* obtido através do software *DrawNetWork* v1.24. com recurso ao filtro EMPOPall_R11.

5.2.2 Haplótipos

A elevada frequência de haplótipos únicos identificados neste estudo, cerca de 71%, é característica de populações Africanas, que apresentam grande heterogeneidade genética. Estudos de ADNmt de outras populações do continente Africano reportam esta elevada percentagem de sequências únicas (Beleza *et al.* 2005; Plaza *et al.* 2004; M. Carvalho *et al.* 2011; Fendt, Röck, *et al.* 2012; Mikkelsen *et al.* 2012), especialmente aquelas que se encontram isoladas, fixas no seu território há muito tempo (Morais 2013).

A identificação de 48 sequências com haplótipos partilhados demonstra a baixa discriminação do estudo do ADNmt em investigações de identificação individual. Existem sequências, que apesar de não pertencerem ao mesmo indivíduo ou à sua linhagem materna, apresentam o mesmo haplótipo, tornando dúbio o seu uso em situações onde a identificação individual é necessária. Neste sentido, a aplicação deste marcador genético na casuística forense depende da frequência do haplótipo na população. Se o haplótipo for frequente, o estudo do ADNmt não é determinante. Por outro lado, na presença de um haplótipo pouco frequente, este marcador genético pode contribuir com informação relevante na resolução de investigações forenses.

Através da base de dados EMPOP, que contem um total de 34 617 sequências, foi possível realizar uma pesquisa por todos os haplótipos aí publicados e averiguar a sua

coincidência com as sequências analisadas neste estudo. Verificou-se que 64 sequências apresentavam haplótipos que não tinham ainda sido descritos na base de dados.

5.2.3 Heteroplasmias

Identificou-se a presença de heteroplasma de comprimento em cerca de 32% das sequências analisadas.

No segmento HVI, o fenómeno de heteroplasma de comprimento está frequentemente associado à transição de uma Timina (T) por uma Citosina (C) na posição 16 189. Esta alteração irá resultar numa extensão de 10 Citosinas, que poderá dar origem a heteroplasma de comprimento, com fragmentos que podem incluir a inserção até 4 Citosinas (C). Este fenómeno origina a sobreposição de picos no eletroferograma e consequente dificuldade de interpretação de resultados a jusante da posição 16193. Identificou-se heteroplasma de comprimento associada a esta transição, em 35 sequências do presente estudo.

No segmento HVII, a heteroplasma de comprimento ocorre entre as posições 302 e 310 onde existe um segmento com 7 Citosinas (C), que pode ser prolongado, perfazendo 11 Citosinas (C). No presente estudo, foram identificadas 27 sequências com heteroplasma de comprimento no segmento HVII.

Em HVIII, a heteroplasma de comprimento é comumente induzida pela inserção de Citosinas (C) entre as posições 568 e 573. Este fenómeno é visível em 4 sequências.

A heteroplasma de posição foi identificada apenas na posição 16183 da amostra ANG163, com a presença de uma Citosina (C) em algumas sequências, e Adenina (A) noutra. Esta mutação pode ter duas origens distintas. Por um lado pode ter sido originada por uma transversão na posição 16183 de uma Adenina (A) por uma Citosina (C), mutação visível em algumas sequências. Por outro lado, pode ter sido criada pela deleção da Adenina (A) da posição 16183 em algumas sequências. Caso se tratasse de uma transição esta heteroplasma de posição não iria originar heteroplasma de comprimento após a mutação. Por outro lado, a presença de uma deleção em apenas algumas sequências iria originar heteroplasma de comprimento a partir da posição 16183. Para confirmar este polimorfismo, seria, por isso, necessário averiguar o eletroferograma e avaliar a possível sobreposição de picos a jusante da posição 16183. O segmento seguinte localizado entre as posições 16184 e 16193 sofre mutação na posição 16189 (transição de uma Timina (T) por uma Citosina (C)) que dá origem a heteroplasma de comprimento associada a esse segmento. Através da técnica utilizada e do eletroferograma obtido não é possível averiguar se a mutação que dá origem à heteroplasma de posição 16183 resulta de uma transição ou de uma deleção.

5.2.4 Haplogrupos

Os macrohaplogrupos M e N e as suas ramificações, persistiram em populações não Africanas após a expansão do Homem moderno a partir de África, sendo atualmente característicos das regiões Euroasiáticas. Por outro lado, o macrohaplogrupo L, que engloba os haplogrupos L0-L6, é limitado geograficamente à região subsariana de África (Salas *et al.* 2004).

A história de África inclui diversas fases de dispersão e migrações de populações que permitiram construir o padrão genético através do continente. A dispersão do povo Bantu representa uma dos maiores movimentos migratórios que ocorreram em África. Esta dispersão terá ocorrido em múltiplas ocasiões e durante diferentes períodos de tempo, traduzindo-se eventualmente, com a fixação do povo Bantu no território Angolano. As populações Khoe-san que aí habitavam terão sido substituídas pelos povos Bantu, ou ter-se-ão movimentado para Sul e Centro do continente Africano, onde, hoje em dia, habitam diversas regiões (incluindo o Sul de Angola).

O haplogrupo L0 (que é considerado uma das primeiras ramificações do ADNmt) surgiu há cerca de 140-160 mil anos e é a subestrutura mais primitiva de pequenas comunidades na região Sul e Este de África (Rosa & Brehem 2011). Este ramo subdivide-se em cinco sub-haplogrupos – L0a, L0b, L0d, L0f e L0k (Salas *et al.* 2002; Salas *et al.* 2004).

Segundo Salas e colaboradores (Salas *et al.* 2002), a classe L0d, a primeira a derivar do L0, surgiu há cerca de 100 mil anos e restringe-se à população Khoe-san no sul de África e às populações da Tanzânia e Angola. Dos 23 haplótipos determinados no presente estudo que pertencem ao ramo L0, apenas 1 se inclui no L0d.

A linhagem L0a teve origem na região oriental do continente Africano há cerca de 40-55 mil anos, sendo comum na região Oriental, Central e Sudoeste do continente; por outro lado, está praticamente ausente na região Norte, Ocidental, e Sul (Salas *et al.* 2002; Rosa & Brehem 2011). Identificaram-se 21 haplótipos pertencentes a L0a.

O haplogrupo L0b é característico da região da Etiópia. Uma das amostras do ramo L0 ficou classificada como L0a'b uma vez que não contem o conjunto de mutações que a permite incluir em L0a ou L0b.

As subclasses L0k e L0f não foram identificadas no presente estudo. O haplogrupo L0k encontra-se quase exclusivamente entre as populações Khoe-san. O ramo L0f é considerado raro e está presente na região Este de África, (Salas *et al.* 2002).

As linhagens L1 surgiram há cerca de 140-150 mil anos e são compostas por L1a, L1b, L1c, L1d, L1e, L1k.

A linhagem L1a é comum nas regiões Este, Central e Sudoeste de África, estando praticamente ausente nas regiões Norte, Oeste e Sul. Muito provavelmente, o ramo L1a terá sido transportado para o Sudeste Africano pelo fluxo Oriental da expansão do povo Bantu,

tendo origem no Este de África (Salas *et al.* 2002). No presente estudo não foi identificada qualquer sequência com haplótipo pertencente ao haplogrupo L1a.

O ramo L1b tem uma distribuição diferente dentro do continente Africano; é predominante na região Ocidental de África, com algum alongamento para Norte e Centro, contudo é raro na região Este, Sudeste e Sul de África. É também comum em populações Afro-Americanas (Salas *et al.* 2002; Rosa & Brehem 2011). Das amostras estudadas, 5 incluem-se neste haplogrupo.

Postula-se que o haplogrupo L1c tenha tido origem no Centro de África, com posterior movimento para a costa atlântica, estando atualmente presente com maior frequência na região Central e Oeste; está presente em populações Bantu do Centro de África, assim como em Afro-americanos. Este haplogrupo inclui 32 amostras, das quais 2 ficam confinadas ao haplogrupo L1c, enquanto 14 pertencem ao L1c1, 5 sequências ao haplogrupo L1c2 e 11 ao haplogrupo L1c3.

O haplogrupo L1d é característico de grupos Khoe-san, com a origem neste povo e o haplogrupo L1e é praticamente restrito ao Este Africano. A linhagem L1k está presente exclusivamente no Sul Africano em populações Khoe-san, sendo, tal como L1d, de origem indígena. Nenhuma das sequências estudadas se inclui nos haplogrupos L1d, L1e e L1k (Salas *et al.* 2002; Rosa & Brehem 2011; Watson *et al.* 1996).

Os haplogrupos L3 e L2 compreendem cerca de 70% da variabilidade materna da região subsariana de África. O haplogrupo L2 compreende as subclasses L2a, L2b, L2c, L2d e L2e.

As linhagens L2a são as mais frequentes e mais difundidas classes de L2 por África, assim como em Afro-Americanos. Os haplogrupo L2b, L2c e L2d parecem estar confinados Oeste e Centro ocidental de África. Dentro do grupo de amostras estudadas, 22 sequências pertencem ao haplogrupo L2a, 4 sequências ao haplogrupo L2b e 6 sequências ao haplogrupo L2c.

As linhagens L3 tiveram origem no Este de África há cerca de 60-75 mil anos. Este haplogrupo está amplamente distribuído pelo continente.

Ambos os haplogrupos L3b e L3d são comuns no Oeste da região subsariana de África e têm também alguma representação em populações Afro-Americanas. L3d contribui com uma elevada percentagem do Sul Africano, sendo mais expressivo em Angola e na Tanzânia. L3b é comum entre a população Bantu.

L3e é a linhagem mais dispersa, frequente e antiga das linhagens Africanas L3, comprimindo aproximadamente um terço de todas as classes L3 na região subsariana de África (Plaza *et al.* 2004) e terá tido origem na região Central do Continente. L3e1 é comum em populações Bantu do Sudeste Africano. L3e2 encontra-se distribuído por África subsariana, com predominância no Sudeste do Continente. Ainda, L3e3 encontra-se essencialmente no Oeste Africano e L3e4 é comum a Este, Centro e Oeste de África.

A difusão do L3f tem uma extensão desde a Etiópia, no Este no continente, até Angola e Moçambique no Sul, bacia do Chade no centro de África, Guiné Bissau no Este e Tunísia no Norte.

Dentro das 55 amostras pertencentes ao haplogrupo L3, 1 delas inclui-se no subhaplogrupo L3a, 5 no subhaplogrupo L3b, 10 no subhaplogrupo L3d, 34 no subhaplogrupo L3e e 5 no subhaplogrupo L3f.

Dentro do haplogrupo L4, apenas foram identificadas 2 sequências, sendo este haplogrupo pequeno e comum na região Este de África.

Apesar de estarem presentes em pequena percentagem, alguns dos haplogrupos identificados nas sequências estudadas não pertencem às linhagens L. Estes não são característicos de populações Africanas, contudo podem estar presentes neste Continente. As populações europeias derivam do haplogrupo N que se ramifica nos haplogrupos H, I, J, K, T, U, V, W e X. Por outro lado as linhagens asiáticas pertencem tanto ao haplogrupo N como ao haplogrupo M.

Dentro do grupo de amostras estudadas, 22 encontravam-se distribuídas entre haplogrupos não Africanos, representadas na Tabela 22.

Tabela 22 – Distribuição dos haplogrupos não Africanos identificados no conjunto de sequências dos imigrantes Angolanos a residir em Lisboa.

Haplogrupo	Número de haplótipos	Predominância geográfica
R0	4	Oeste Euroásia
H	5	Oeste Euroásia
M1	1	Oeste Euroásia
J	1	Oeste Euroásia
K	3	Oeste Euroásia
T1	1	Oeste Euroásia
T2	4	Oeste Euroásia
U5	3	Oeste Euroásia

A baixa contribuição de haplogrupos europeus na população Angolana a residir em Lisboa, confirma a informação histórica referente o período de colonização de Angola. Este país esteve sob domínio português desde o século XV e, durante esta fase da colonização, indivíduos portugueses e, primeiramente, do sexo masculino, migraram para a colónia portuguesa. Em possíveis miscigenações no território Angolano durante esta época, os indivíduos europeus não contribuíram com o seu genoma mitocondrial. A contribuição genética feminina terá sido, maioritariamente, subsariana. Esta informação confirma a mistura entre homens europeus e mulheres Africanas, algo comum noutras colónias portuguesas (Trovada

et al. 2004; Pereira *et al.* 2001; Gonçalves *et al.* 2003). Por outro lado, a presença dos imigrantes Angolanos em território português é relativamente recente, uma vez que o seu movimento migratório teve início por volta das décadas de 80/90. Neste sentido, a contribuição do genoma mitocondrial europeu neste grupo de indivíduos, resultante de possíveis relações inter raciais sucedidas em Portugal, é reduzida.

A elevada percentagem de linhagens maternas pertencentes ao macrohaplogrupo L no conjunto de indivíduos estudado, determina que a presença deste haplogrupo em Portugal se deve ao recente fluxo migratório de indivíduos oriundos de ex-colónias portuguesas, e não à evolução do macrohaplogrupo L no território europeu. Esta grande contribuição de haplogrupos característicos de regiões subsarianas de África no grupo de indivíduos estudados demonstra também a elevada variabilidade genética introduzida pela população imigrante Angolana em Lisboa. Os recentes fluxos migratórios irão provocar uma alteração no padrão genético mitocondrial da população residente em Lisboa. Num futuro próximo, será expectável a presença de indivíduos naturais de Lisboa, com haplótipos característicos de populações Africanas. Desta forma, sublinha-se a relevância do presente estudo na atualização da informação genética da população residente em Lisboa. A informação obtida com o acompanhamento do fluxo migratório e com o estudo genético das populações imigrantes, pode ser útil em investigações de casuística forense e de genética populacional.

A ausência de haplogrupos pertencentes às populações Khoe-san, nomeadamente L0d e L0k, assim como a presença de haplogrupos característico de populações Bantu no grupo de amostras estudadas, confirmam a elevada ou mesmo total substituição do povo Khoe-san pelas linhagens Bantu durante o seu fluxo migratório em África. Esta informação é confirmada por outros autores (Alves-Silva *et al.* 2000; Plaza *et al.* 2004).

Alguns dos haplogrupos determinados na população em estudo, são também descritos e característicos de populações Afro-americanas. Estas informações confirmam os relatos históricos do transporte de escravos Angolanos para a América, com a contribuição da sua informação genética na formação de povos americanos.

5.2.5 Parâmetros de diversidade

Os valores de diversidade de sequência e de diversidade nucleotídica foram calculados para a região controlo total do ADNmt da população em estudo, sendo $\hat{H} = 0,996$ e $\pi = 0,016$. Estes resultados foram posteriormente comparados com os publicados por Batini (Batini *et al.* 2011) e Morais (Morais 2013), também referentes à região controlo total do ADNmt. Foi possível averiguar uma semelhança entre os valores de diversidade de sequência da população em estudo e os valores de diversidade de sequência das restantes populações

consideradas. O mesmo acontece com o valor de diversidade nucleotídica, sendo o valor obtido para a população em estudo ligeiramente superior aos restantes.

Para as regiões HVI e HVII do ADNmt foi possível obter resultados publicados referentes ao índice de diversidade de sequência para a população Bantu em Angola (Plaza *et al.* 2004) e para a população de Moçambique (Pereira *et al.* 2001). O índice de diversidade de sequência obtido para os indivíduos dos povos Bantu em Angola é muito próximo do valor de diversidade de sequência dos imigrantes Angolanos a residir em Lisboa.

Os parâmetros de diversidade foram também calculados para o segmento HVI da região controlo do ADNmt. O valor de diversidade de sequência referente a esta região é elevado, sendo que o mesmo é visível nas populações Bantu publicadas. O valor de diversidade nucleotídica para a região HVI também se encontra próximo dos descritos na literatura para as outras populações Africanas.

As populações Africanas apresentam maior variabilidade genética quando comparadas com populações de outras origens. A elevada variabilidade genética da população em estudo, demonstrada pelos valores de diversidade de sequência e diversidade nucleotídica calculados, pode ser explicada pela conservação do padrão genético mitocondrial Africano dos indivíduos Angolanos a residir em Lisboa.

Como seria de esperar, o valor de diversidade de sequência é superior quando calculado com a totalidade da região controlo do ADNmt do que apenas com o segmento HVI, pelo aumento do número de nucleótidos considerados no estudo. Ainda, o valor de diversidade de sequência é superior quando efetuado manualmente (onde são consideradas as inserções) do que quando é calculado pelo *software* Arlequin (onde as inserções não são tidas em consideração).

5.2.6 Significado filogenético

No estudo filogenético, os haplótipos de ADNmt dos 173 indivíduos Angolanos a residir em Lisboa, foram comparadas com haplótipos de ADNmt pertencentes a indivíduos de outras populações Africanas publicadas. São poucas as populações Africanas com sequências de ADNmt que incluem a totalidade da região controlo da molécula. Neste sentido apenas foram consideradas para comparação filogenética as populações do Gana (Fendt, Röck, *et al.* 2012), da Somália (Mikkelsen *et al.* 2012), de indígenas Khoe-san (Fendt, Huber, *et al.* 2012) e de imigrantes Cabo-Verdianos a residir em Lisboa (Morais 2013).

Através da matriz de distâncias interpopulacionais determinou-se que a população de imigrantes Angolanos a residir em Lisboa encontra-se geneticamente distante da população Khoe-san ($F_{st}=0.24677$; $p=0.00000$). Outros estudos confirmam esta fraca ou mesmo ausência de contribuição do povo Khoe-san no padrão genético Angolano (Plaza *et al.* 2004; Salas *et al.* 2004; Alves-Silva *et al.* 2000). Por outro lado, a população em estudo encontra-se

geneticamente próxima da população do Gana ($F_{st}=0.02746$; $p=0.00000$), seguida da Somália ($F_{st}=0.03588$; $p=0.00000$), sendo por isso mais distante da população de imigrantes, oriundos de Cabo Verde, a residir em Lisboa ($F_{st}=0.04408$; $p=0.00000$) (Tabela 18).

Com a ocupação do território do Gana pelos colonizadores portugueses no ano de 1471, iniciou-se a comercialização de diversas mercadorias e escravos entre o Gana e outros países Africanos (Page 2005). Fendt e colaboradores, no ano de 2012, estudaram a totalidade da região controlo do ADNmt de 191 indivíduos do Gana. A grande maioria das sequências estudadas, 98.4%, pertencem aos haplogrupos L1-L3, característicos de África e comuns nas regiões Oestes do Continente. A troca de escravos entre Angola e Gana pode explicar a proximidade entre ambas as populações.

No ano de 2012, Mikkelsen e colaboradores estudaram o ADNmt de 190 indivíduos da Somália. Dos haplótipos identificados, 60.5%, pertencem aos haplogrupos Africanos L1-L4 e L5. Os restantes haplótipos (39.5%) foram atribuídos aos haplogrupos M1, N1 e R. Alguns dos haplogrupos estudados por estes autores foram também identificados no grupo de amostras de imigrantes Angolanos.

Morais, em 2013, estudou 103 amostras de indivíduos Cabo-verdianos a residir em Lisboa. A distância genética que a população Angolana imigrante a residir em Lisboa apresenta relativamente aos imigrantes Cabo Verdianos em Lisboa, pode estar relacionada com o facto de os imigrantes de Cabo Verde apresentarem a estrutura genética do seu país de origem. Cabo Verde é um país composto por um conjunto de ilhas isoladas o que, durante períodos históricos, limitou a sua comunicação com outros países do continente Africano. Neste sentido, apesar de os imigrantes Angolanos também apresentarem um padrão genético característico de populações Africanas, o isolamento de Cabo Verde, explica a distância genética existente entre os imigrantes oriundos dos dois países Africanos.

Valores de p iguais ou superiores a 0.05 indicam a existência de semelhanças entre populações. Os valores de p obtidos entre a população Angolana a residir em Lisboa e as restantes populações consideradas para a região controlo total do ADNmt ($p = 0.00$), indicam que a população em estudo apresenta diferenças significativas com os restantes grupos populacionais Africanos estudados.

Para fins comparativos e uma vez que não existem dados relativos à população Angolana, a residir em Angola e que incluam a totalidade da região controlo do ADNmt, as sequências estudadas foram reduzidas aos segmentos HVI e HVII. Desta forma foi possível realizar estudos comparativos dos imigrantes Angolanos a residir em Lisboa com diversos grupos étnico linguísticos pertencentes ao povo Bantu (Coelho *et al.* 2009), com população Angolana a residir em Angola (Carvalho *et al.* 2006; Beleza *et al.* 2005), e com outras populações de países Africanos: Moçambique (Pereira *et al.* 2001), Guiné Bissau (M Carvalho *et al.* 2011) e São Tomé e Príncipe (Trovoada *et al.* 2004). As populações anteriormente

estudadas com a região controlo total do ADNmt foram também reduzidas a HVI e HVII e foram introduzidas nesta fase do estudo.

Verificou-se uma proximidade genética entre a população imigrante Angolana a residir em Lisboa e as populações de Angola, de grupos étnicos pertencentes ao povo Bantu e de São Tomé e Príncipe.

A reduzida distância genética apresentada entre a população em estudo e a população Angolana a residir em Angola seria expectável, indicando o baixo contributo genético materno de origem europeia e a elevada contribuição de indivíduos Angolanos no grupo de amostras estudadas. Desta forma, a população Angolana a residir em Lisboa apresenta ainda o padrão genético mitocondrial do seu território de origem. De igual forma, verifica-se uma reduzida distância entre a população em estudo e os diversos grupos étnico-linguísticos Bantu, sendo possível averiguar que alguns deles, têm uma maior influência na constituição do povo Angolano, enquanto outros, apresentam uma menor contribuição. O grupo étnico-linguístico Ovimbundo, que representa cerca de 37% da constituição de Angola, é o que se encontra mais próximo da população imigrante Angolana. Contrariamente, o grupo étnico-linguístico Kuvale, encontra-se mais distante. Verificou-se novamente a elevada distância da população em estudo com o povo Khoe-san. Estes resultados evidenciam a elevada taxa de substituição do povo Khoe-san pelas linhagens Bantu no território Angolano, sendo o povo Bantu o grande contribuidor do padrão genético materno da população Angolana (Plaza *et al.* 2004).

O arquipélago de São Tomé e Príncipe situa-se na costa Ocidental de África. Este país tornou-se colónia Portuguesa no século XV. Até 1876, ano em que Portugal aboliu o tráfico de escravos, este país era um ponto de comercialização de escravos para as restantes colónias portuguesas criadas no continente Africano e no continente americano. Estes relatos históricos justificam a sua proximidade com os imigrantes Angolanos em Lisboa.

Na comparação entre a população de imigrantes Angolanos a residir em Lisboa e as populações de São Tomé e Príncipe, Angola e vários grupos étnico-linguísticos Bantu (Ganguela, Nhaneca-Humbe e Ovimbundo) verifica-se a presença de valores de *p* superiores a 0.05 o que, mais uma vez, confirma a semelhança e proximidade entre estas populações.

6. Conclusões

- O número de imigrantes em Portugal aumentou consideravelmente a partir dos anos 70, fenómeno relacionado com a fase pós-colonial de muitos países de língua portuguesa. Angola revelou-se um dos países mais representativos como origem de imigrantes em Portugal, com uma predominância na região de Lisboa.

- Angola é um país localizado na costa Atlântica Sul do Continente Africano, tendo sido habitado inicialmente pelo povo Khoe-san. Mais tarde estes foram substituídos e/ou expulsos pela população Bantu, oriunda da região Central/Oeste de África. No século XV Angola foi ocupada pelos colonizadores portugueses. Estes dois acontecimentos históricos são considerados os principais modeladores genéticos do território Angolano.

- Para avaliar a contribuição genética da população imigrante Angolana a residir em Lisboa, analisou-se a região controlo do ADNmt de 173 indivíduos de ascendência Angolana a habitar na região de Lisboa.

- Amplificou-se a totalidade da região controlo do ADNmt, numa extensão de 1122pb, através da utilização de dois pares de *primers* - L15997/H016 e L16555/H639. A sequenciação dos segmentos obtidos foi realizada através do método de sequenciação de Sanger. A metodologia utilizada revelou ser eficiente para a realização do objetivo proposto, permitindo a obtenção de sequências referentes à totalidade da região controlo do ADNmt, com elevada qualidade.

- O uso da totalidade da região controlo do ADNmt revelou ser de eficiência superior, comparativamente a estudos anteriores da molécula que apenas incluíam a sequenciação dos segmentos HVI e HVII ou do segmento HVI isolado da região controlo do ADNmt. Os polimorfismos presentes na região controlo total do ADNmt ajudam a clarificar os haplótipos e a discriminar os haplogrupos a que as sequências se inserem.

- Os haplótipos determinados no conjunto de amostras estudadas pertencem, em grande parte, a haplogrupos característicos da região subsariana de África. Apenas uma pequena percentagem (≈13%) de haplogrupos é característica de populações euroasiáticas. É possível confirmar a baixa influência europeia na constituição do padrão genético de ADNmt do indivíduo Angolano e, conseqüentemente, no padrão genético dos imigrantes Angolanos a residir em Lisboa.

- O elevado número de posições polimórficas e respetivos polimorfismos, confirma dados anteriores relativos à elevada heterogeneidade genética presente em populações do continente Africano. A grande percentagem de haplótipos únicos presente do estudo realizado é também característica de populações Africanas.

- Através da obtenção das distâncias genéticas entre a população do imigrantes Angolanos a residir em Lisboa com outras populações Africanas, confirmou-se a proximidade deste povo imigrante com os indivíduos Angolanos a residir em Angola e com diversas populações Bantu.

- A análise do ADNmt dos indivíduos Angolanos a residir em Lisboa, reforça a história da criação do povo Angolano, estando em concordância com relatos arqueológicos e linguísticos. Confirma-se, desta forma, que o padrão genético do ADNmt de Angola e, consequentemente de imigrantes Angolanos a residir em Lisboa resulta, em grande parte, da expansão do povo Bantu, que com a sua chegada a este território, expulsou os primeiros habitantes da região, o povo Khoe-san. Por outro lado, colonização de Angola terá ocorrido maioritariamente por indivíduos europeus do sexo masculino, que durante possíveis miscigenações com indígenas do território, não terão contribuído com o seu material genético mitocondrial. Ainda, a presença de imigrantes Angolanos em Portugal é recente, sendo que ainda não decorreu tempo suficiente para que as miscigenações aí desenvolvidas tenham considerável contribuição europeia.

- O presente estudo revela a importância da análise de ADNmt em investigações populacionais. É possível, com base no estudo do ADNmt, confirmar relatos históricos relativos a trajetos migratórios e avaliar a contribuição genética das populações migratórias no território de fixação.

- Tendo em consideração o número crescente de imigrantes Angolanos em Portugal, a informação retirada a partir do estudo do ADNmt do grupo de indivíduos Angolanos a residir em Lisboa realça a importância destes estudos na caracterização genética dos indivíduos a residir em Lisboa. Num futuro próximo, indivíduos naturais e nacionais de Lisboa, irão apresentar haplótipos, até então considerados como tipicamente Africanos.

- Informação adicional, retirada partir do estudo da região codificante do ADNmt, iria permitir retirar conclusões mais detalhadas dos haplótipos de ADNmt dos indivíduos e, desta forma, da constituição genética da população Angolana imigrante a residir em Lisboa.

7. Bibliografia

- Afonso Costa, H. et al., 2010. Mitochondrial DNA sequence analysis of a native Bolivian population. *Journal of forensic and legal medicine*, 17(5), pp.247–53. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20569950> [Accessed October 21, 2013].
- Alberts, B. et al., 2010. *Biologia Molecular da célula* 5th ed., Ediciones Omega, S.A.
- Alonso, A. et al., 2002. Results of the 1999-2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor. *Forensic science international*, 125(1), pp.1–7.
- Alvarez, J.C. et al., 2007. Characterization of human control region sequences for Spanish individuals in a forensic mtDNA data set. *Legal medicine (Tokyo, Japan)*, 9(6), pp.293–304. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17616421> [Accessed October 21, 2013].
- Alves-Silva, J. et al., 2000. The ancestry of Brazilian mtDNA lineages. *American journal of human genetics*, 67(2), pp.444–61. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1287189&tool=pmcentrez&rendertype=abstract>.
- Anderson, S. et al., 1981. Sequence and organization of the human mitochondrial genome. *Nature*, 290, pp.457–465.
- Andrews, R.M. et al., 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature genetics*, 23, p.147.
- Anon, 1975. *História de Angola*, Afrontamento.
- Anon, 1992. *Iniciação à história da Angola*, República Popular de Angola, Ministério da Educação.
- Anon, 2014. PORDATA: Base de Dados Portugal Contemporâneo. Available at: www.pordata.pt [Accessed February 8, 2014].
- Awadalla, P., Eyre-Walker, A. & Smith, J.M., 1999. Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA. *Science*, 286, pp.2524–2525.
- Bandelt, H.-J., van Oven, M. & Salas, A., 2012. Haplogrouping mitochondrial DNA sequences in Legal Medicine/Forensic Genetics. *International journal of legal medicine*, 126(6), pp.901–16. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22940763> [Accessed August 14, 2014].
- Batini, C. et al., 2011. Insights into the demographic history of African Pygmies from complete mitochondrial genomes. *Molecular biology and evolution*, 28(2), pp.1099–110. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21041797> [Accessed August 28, 2014].

- Beleza, S. et al., 2005. The genetic legacy of western Bantu migrations. *Human genetics*, 117(4), pp.366–75. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15928903> [Accessed August 9, 2014].
- Brandstätter, A. & Parson, W., 2003. Mitochondrial DNA heteroplasmy or artefacts - a matter of the amplification strategy? *International journal of legal medicine*, 117(3), pp.180–4. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12799738> [Accessed October 21, 2013].
- Brown, T.A., 2006. *Genomes* 3 3rd ed., Garland Science.
- Brown, W.M., 1980. Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. , 77(6), pp.3605–3609.
- Budowle, B. et al., 1999. Mitochondrial DNA regions HVI and HVII population data. *Forensic Science International*, 103(1), pp.23–25.
- Butler, J.M., 2011. *Advanced Topics in Forensic DNA Typing: Methodology*, Elsevier.
- Butler, J.M., 2005. *Forensic DNA Typing*, Elsevier Academic Press.
- Calloway, C.D. et al., 2000. The frequency of heteroplasmy in the HVII region of mtDNA differs across tissues types and increase with age. *American journal of human genetics*, 66(4), pp.13384–1397.
- De Camargo, M.A. et al., 2011. No relationship found between point heteroplasmy in mitochondrial DNA control region and age range, sex and haplogroup in human hairs. *Molecular biology reports*, 38(2), pp.1219–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20563855> [Accessed March 25, 2014].
- Cann, R.L., Stoneking, M. & Wilson, A.C., 1987. Mitochondrial DNA and human evolution. *Nature*, 325, pp.31–36.
- Carracedo, A. et al., 2000. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. *Forensic science international*, 110, pp.17–85. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11054055>.
- Carvalho, M. et al., 2011. Mitochondrial control region data of 3 ethnic groups from angola. *Forensic science international. Genetics*, 3, pp.e125–e126.
- Carvalho, M. et al., 2006. Mitochondrial DNA Analysis in a Population from Angola.
- Carvalho, M. et al., 2003. mtDNA analysis in Portuguese populations (Central Portugal and Azores Islands): polymorphic sites in control region sequences. *International Congress Series*, 1239, pp.535–539. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0531513102005800>.
- Carvalho, M et al., 2011. Paternal and maternal lineages in Guinea-Bissau population. *Forensic science international. Genetics*, 5(2), pp.114–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21051306> [Accessed July 29, 2014].
- Castri, L. et al., 2009. mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in sub-Saharan Africa. *American journal of physical anthropology*, 140(2), pp.302–11. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19425093> [Accessed August 8, 2014].

- Castro de Guerra, D. et al., 2012. Sequence variation of mitochondrial DNA control region in North Central Venezuela. *Forensic science international. Genetics*, 6(5), pp.e131–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22189355> [Accessed October 21, 2013].
- Coelho, M. et al., 2009. On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola. *BMC evolutionary biology*, 9, p.80. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2682489&tool=pmcentrez&rendertype=abstract> [Accessed July 24, 2014].
- Comas, D., Paabo, S. & Bertranpetit, J., 1995. Heteroplasmy in the control region of human mitochondrial DNA. *Genome Research*, 5(1), pp.89–90. Available at: <http://www.genome.org/cgi/doi/10.1101/gr.5.1.89> [Accessed August 14, 2014].
- Cummins, J., 1998. Mitochondrial DNA in mammalian reproduction. *Reviews of reproduction*, 3(3), pp.172–82. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9829552>.
- Davis, B.C.L., 1998. *Mitochondrial DNA : State of Tennessee v . Paul Ware*,
- Erlich, H.A., 1989. Polymerase chain reaction. *Journal of Clinical Immunology*, 9(6), pp.437–447. Available at: <http://link.springer.com/10.1007/BF00918012>.
- Excoffier, L., Laval, G. & Schneider, S., 2005. Arlequin (version 3.0); an integrated software package for population genetics data analysis. *Evol Bionform Online*, 1, pp.47–50.
- Falkenberg, M., Larsson, N.-G. & Gustafsson, C.M., 2007. DNA replication and transcription in mammalian mitochondria. *Annual review of biochemistry*, 76, pp.679–99. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/17408359> [Accessed May 25, 2014].
- Felsenstein, J., 1989. PHYLIP: phylogeny interface package (version 3.2). *Cladistics*, 5, pp.164–166.
- Fendt, L., Huber, G., et al., 2012. Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages. *Forensic science international. Genetics*, 6(5), pp.662–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22436123> [Accessed July 29, 2014].
- Fendt, L., Röck, A., et al., 2012. MtDNA diversity of Ghana: a forensic and phylogeographic view. *Forensic science international. Genetics*, 6(2), pp.244–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3314991&tool=pmcentrez&rendertype=abstract> [Accessed July 29, 2014].
- Figueiredo, J. de, 1945. *Apointamentos sobre história de Angola*, Livraria Lello.
- Gonçalves, R. et al., 2003. Y-chromosome lineages in Cabo Verde Islands witness the diverse geographic origin of its first male settlers. *Human genetics*, 113(6), pp.467–72. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/12942365> [Accessed October 21, 2013].
- Gonder, M.K. et al., 2006. Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. *Oxford journal*.
- Gravel, S. et al., 2011. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29), pp.11983–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3142009&tool=pmcentrez&rendertype=abstract> [Accessed August 24, 2014].

- Hagelberg, E. et al., 1999. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. *Proceedings of the royal society*, 266, pp.485–492.
- Hartl, D.L. & Clark, A.G., 1997. *Principles Population Genetics* 3rd ed., Sinauer Associates, Inc., Publishers.
- Haviland, W.A. et al., 2012. *Essence of Anthropology* 3rd ed., Wadsworth.
- Henn, B.M., Cavalli-Sforza, L.L. & Feldman, M.W., 2012. The great human expansion. *Proceedings of the National Academy of Sciences of the United States of America*, 109(44), pp.17758–64. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3497766&tool=pmcentrez&rendertype=abstract> [Accessed December 17, 2013].
- Holland, M.M. & Parsons, T.J., 1999. Mitochondrial DNA Sequence Analysis - Validation and Use for Forensic Casework. *Forensic science review*, 11.
- Howell, N., Kubacka, I. & Mackey, D. a, 1996. How rapidly does the human mitochondrial genome evolve? *American journal of human genetics*, 59(3), pp.501–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1914922&tool=pmcentrez&rendertype=abstract>.
- Howell, N. & Smejkal, C.B., 2000. Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction. *American journal of human genetics*, 66(5), pp.1589–98. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1378018&tool=pmcentrez&rendertype=abstract>.
- Iborra, F.J., Kimura, H. & Cook, P.R., 2004. The functional organization of mitochondrial genomes in human cells. *BMC biology*, 14, pp.1–14.
- Jazin, E.E. et al., 1996. Human brain contains high levels of heteroplasmy in the noncoding regions of mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 93(22), pp.12382–7. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=38000&tool=pmcentrez&rendertype=abstract>.
- Johnson, A.A. & Johnson, K.A., 2001. Fidelity of nucleotide incorporation by human mitochondrial DNA polymerase. *The Journal of biological chemistry*, 276(41), pp.38090–6. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11477093> [Accessed September 23, 2014].
- K, M. et al., 1986. Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction. *Cold Spring Harbor Symposia on Quantitative Biology*, pp.263–273.
- Knapp, M. et al., 2012. Complete mitochondrial DNA genome sequences from the first New Zealanders. *Proceedings of the National Academy of Sciences of the United States of America*, 109(45), pp.18350–4. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3494939&tool=pmcentrez&rendertype=abstract> [Accessed May 30, 2014].
- Laval, G. et al., 2010. Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PloS one*, 5(4), p.e10284. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2858654&tool=pmcentrez&rendertype=abstract> [Accessed August 30, 2014].

- Lewin, B., 2008. *Genes IX*, ARTMED.
- Lima, G. et al., 2006. HVI and HVII sequence polymorphisms of the human mtDNA in the North of Portugal: Population data and maternal lineages. *International Congress Series*, 1288, pp.133–135. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0531513105015803> [Accessed October 21, 2013].
- Lodish, H. et al., 2000. *Molecular Cell Biology*, New York, NY: W. H. Freeman.
- Lutz, S. et al., 1996. mtDNA as a tool for identification of human remains. Identification using mtDNA. *International journal of legal medicine*, 109(4), pp.205–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9007636>.
- Maca-Meyer, N. et al., 2001. Major genomic mitochondrial lineages delineate early human expansions. *BMC genetics*, 2, p.13. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=55343&tool=pmcentrez&rendertype=abstract>.
- Mairal, Q. et al., 2013. Linguistic isolates in Portugal: insights from the mitochondrial DNA pattern. *Forensic science international. Genetics*, 7(6), pp.618–23. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/24041913> [Accessed January 7, 2014].
- Makaowski, W., 2001. The human genome structure and organization. *Acta Biochimica Polonica*, 48(3).
- Malyarchuk, B.A. et al., 2002. Analysis of phylogenetically reconstructed mutational spectra in human mitochondrial DNA control region. *Human genetics*, 111(1), pp.46–53.
- Martin, W.F. & Muller, M., 2007. *Origin of Mitochondria and Hydrogenosomes*, Springer.
- McEvoy, B.P. et al., 2011. Human population dispersal “Out of Africa” estimated from linkage disequilibrium and allele frequencies of SNPs. *Genome research*, 21(6), pp.821–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3106315&tool=pmcentrez&rendertype=abstract> [Accessed August 30, 2014].
- Mendes, M.F., Rebelo dos Santos, J. & Rego, C., 2011. Imigrantes Angolanos em Portugal: breve caracterização e contributos para a dinâmica populacional. *XI CONLAB*.
- Micklos, D.A. & Freyer, G.A., 2003. *DNA Science: A First Course*, CSHL Press.
- Mikkelsen, M. et al., 2012. Forensic and phylogeographic characterisation of mtDNA lineages from Somalia. *International journal of legal medicine*, 126(4), pp.573–9. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22527188> [Accessed July 29, 2014].
- Morais, P., 2013. *Estudo do DNA mitocondrial de um grupo de imigrantes oriundos de Cabo Verde residentes em Lisboa Mestrado em Biologia Molecular em Saúde*.
- Nagai, A. et al., 2003. Sequence polymorphism of mitochondrial DNA in Japanese individuals from Gifu Prefecture. , 5, pp.210–213.
- Nass, M.M.K. & Nass, S., 1963. Intramitochondrial fiber with DNA characteristics. *The journal of cell biology*, 19.
- Nei, M. & Tajima, F., 1981. Genetic drift and estimation of effective population size. , pp.625–640.

- Nelson, D.L. & Cox, M.M., 2004. *Lehninger Principles of Biochemistry* 4th ed., W H Freeman & Co.
- Newman, J.L., 1997. *The Peopling of Africa: A Geographic Interpretation*, Yale University Press;
- Oppenheimer, S., 2012. A single southern exit of modern humans from Africa: Before or after Toba? *Quaternary International*, 258, pp.88–89.
- Van Oven, M. & Kayser, M., 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. *Human mutation*, 30(2), pp.E386–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18853457> [Accessed October 21, 2013].
- Page, R.D.M., 1996. Treeview: an application to display phylogenetic trees on personal computers. *Comput Appl Biosc*, 12(4), pp.357–358.
- Page, W.F., 2005. *Encyclopedia Of African History And Culture*, 5 Vol.,
- Pakendorf, B. & Stoneking, M., 2005. Mitochondrial DNA and human evolution. *Annual review of genomics and human genetics*, 6, pp.165–83. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/16124858> [Accessed October 18, 2013].
- Parson, W. et al., 2014. DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic science international. Genetics*, 13C, pp.134–142. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/25117402> [Accessed August 21, 2014].
- Parson, W. et al., 1998. Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *International journal of legal medicine*, 111(3), pp.124–32. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9587794>.
- Parson, W. & Dür, A., 2007. EMPOP-a forensic mtDNA database. *Forensic science international. Genetics*, 1(2), pp.88–92. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/19083735> [Accessed October 21, 2013].
- Pereira, L. et al., 2001. Prehistoric and historic traces in the mtDNA of Mozambique : insights into the Bantu expansions and the slave trade. *Ann. Hum. Genet.*, 65, pp.439–458.
- Pereira, L., Prata, M.J. & Amorim, a, 2000. Diversity of mtDNA lineages in Portugal: not a genetic edge of European variation. *Annals of human genetics*, 64(Pt 6), pp.491–506. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11281213>.
- Phillipson, D.W., 2005. *African Archaeology* 3 edition., Cambridge University Press.
- Pierce, B.A., 2012. *Genetics: A Conceptual Approach* 4th ed., W. H. Freeman.
- Pinheiro, M. de F.T., 2010. *Genética Forense: Perspectivas da identificação genética*, Porto: edições Universidade Fernando Pessoa.
- Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, Á., et al., 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human genetics*, pp.439–447.
- Plaza, S., Salas, A., Calafell, F., Corte-Real, F., Bertranpetit, J., Carracedo, A., et al., 2004. Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola. *Human genetics*, 115(5), pp.439–47. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15340834> [Accessed August 6, 2014].

- Raeymaeker, J. de, 2012. *À descoberta de Angola* 1st ed., Oficina do Livro - Sociedade Editorial Lda.
- Rapley, R. & Whitehouse, D., 2007. *Molecular Forensics*, Wiley.
- Redinha, J., 2009. *Etnias e culturas de Angola*,
- Regateiro, F. de J., 2007. *Manual de genética Médica*, Imprensa da Universidade de Coimbra.
- Relethford, J.H., 2012. *Human Population Genetics*, Wiley-Blackwell.
- Roberts, K.A. & Calloway, C., 2011. Characterization of mitochondrial DNA sequence heteroplasmy in blood tissue and hair as a function of hair morphology. *Journal of forensic sciences*, 56(1), pp.46–60. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/20840293> [Accessed October 21, 2013].
- Rosa, A. & Brehem, A., 2011. African human mtDNA phylogeography at-a-glance. *Journal of anthropological sciences = Rivista di antropologia : JASS / Istituto italiano di antropologia*, 89, pp.25–58. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21368343> [Accessed December 17, 2013].
- Ruiz-Pesini, E. et al., 2007. An enhanced MITOMAP with a global mtDNA mutational phylogeny. *Nucleic acids research*, 35(Database issue), pp.D823–8. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1781213&tool=pmcentrez&rendertype=abstract> [Accessed June 7, 2014].
- Salas, A. et al., 2004. The African diaspora: mitochondrial DNA and the Atlantic slave trade. *American journal of human genetics*, 74(3), pp.454–65. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1182259&tool=pmcentrez&rendertype=abstract>.
- Salas, A. et al., 2002. The making of the African mtDNA landscape. *American journal of human genetics*, 71(5), pp.1082–1111.
- Samuels, D.C. et al., 2013. Recurrent tissue-specific mtDNA mutations are common in humans. *PLoS genetics*, 9(11), p.e1003929. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3820769&tool=pmcentrez&rendertype=abstract> [Accessed August 30, 2014].
- Sanger, F. & Nicklen, S., 1977. DNA sequencing with chain-terminating inhibitors. , 74(12), pp.5463–5467.
- Scheible, M. et al., 2011. Mitochondrial DNA control region variation in a Kuwaiti population sample. *Forensic science international. Genetics*, 5(4), pp.e112–3. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/21555259> [Accessed October 21, 2013].
- Schwartz, Ma. & Vissing, J., 2002. Paternal Inheritance of Mitochondrial DNA. *The New England Journal of Medicine*, 347(8), pp.576–580.
- Scientific Working Group for DNA Analysis Methods, 2003. Guidelines for mitochondrial DNA (mtDNA) Nucleotide sequence information. *Forensic science communications*, 5.
- Scientific Working Group on DNA Analysis Methods, 2013. Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories. , pp.1–23.
- Serviço de Estrangeiros e Fronteiras, 2013. *Relatório de Imigração Fronteiras e Asilo*,

- Shoubridge, E. a, 2000. Mitochondrial DNA segregation in the developing embryo. *Human reproduction (Oxford, England)*, 15 Suppl 2, pp.229–34. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11041528>.
- Sondheimer, N. et al., 2011. Neutral mitochondrial heteroplasmy and the influence of aging. *Human molecular genetics*, 20(8), pp.1653–9. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3063991&tool=pmcentrez&rendertype=abstract> [Accessed March 25, 2014].
- Stead, M. & Rorison, S., 2009. *Angola*, The Globe Pequot Press Inc.
- Stewart, J.R. & Stringer, C.B., 2012. Human evolution out of Africa: the role of refugia and climate change. *Science (New York, N.Y.)*, 335(6074), pp.1317–21. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/22422974> [Accessed December 11, 2013].
- Sutovsky, P. et al., 2000. Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. *Biology of reproduction*, 63(2), pp.582–90. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10906068>.
- Tattersall, I., 2009. Human origins: out of Africa. *Proceedings of the National Academy of Science of the United States of America*, 106(38), pp.16018–16021. Available at: <http://www.pnas.org/content/106/38/16018.short> [Accessed March 14, 2014].
- Taylor, R.W. & Turnbull, D.M., 2007. Mitochondrial DNA mutations in human disease. , 6(5), pp.389–402.
- Templeton, A., 2002. Out of Africa again and again. *Nature*, 416(6876), pp.45–51. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1949410&tool=pmcentrez&rendertype=abstract>.
- Thomas, H., 1999. *The slave trade: the story of the Atlantic slave trade*, Simon & Schuster.
- Trovoada, M.J. et al., 2004. Insights from pattern of mtDNA variation into the genetic history of São Tomé e Príncipe. *International Congress Series*, 1261, pp.377–379. Available at: <http://linkinghub.elsevier.com/retrieve/pii/S0531513103016339> [Accessed July 30, 2014].
- Tully, G. et al., 2001. Considerations by the European DNA profiling (EDNAP) group on the working practices , nomenclature and interpretation of mitochondrial DNA profiles. *Forensic science international*, 124, pp.83–91.
- Underhill, P. a & Kivisild, T., 2007. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual review of genetics*, 41, pp.539–64. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/18076332> [Accessed October 21, 2013].
- Voet, D., Voet, J.G. & Pratt, C.W., 2012. *Fundamental of Biochemistry* 4th ed., Wiley.
- Walsh, P.S., Metzger, D.A. & Higuchi, R., 1991. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *BioTechniques*, 10, pp.506–513.
- Watson, E. et al., 1996. mtDNA sequence diversity in Africa. *American journal of human genetics*, 59(2), pp.437–44. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1914737&tool=pmcentrez&rendertype=abstract>.
- Wilson, M.R. et al., 1995. Validation of mitochondrial DNA sequencing for forensic casework analysis. *International journal of legal medicine*, 108(2), pp.68–74. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/8547161>.

Wolpoff, M.H., Hawks, J. & Caspari, R., 2000. Multiregional, not multiple origins. *American journal of physical anthropology*, 112(1), pp.129–36. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10766948>.

Zau, F., 2002. *Angola: Trilhos para o Desenvolvimento*, Universidade Aberta.

Zimmermann, B., Röck, A.W. & Parson, W., 2012. Drawing and interpreting quasi-median networks with A short introduction.

8. Anexos



Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen



Genetic portrait of Lisboa immigrant population from Angola with mitochondrial DNA

Filipa Simão^{a,b}, Heloísa Afonso Costa^{a,f}, Cláudia Vieira da Silva^{a,f}, Teresa Ribeiro^{a,f},
Maria João Porto^{a,f}, Jorge Costa Santos^{a,c,d,f}, António Amorim^{a,d,e,f,*}

^aInstituto Nacional de Medicina Legal e Ciências Forenses, Portugal

^bFaculdade de Ciências e Tecnologias da Universidade Nova de Lisboa, Portugal

^cFaculdade de Medicina da Universidade de Lisboa, Portugal

^dInstituto Superior de Ciências de Saúde Egas Moniz, Portugal

^eFaculdade de Ciências da Universidade de Lisboa, Portugal

^fCENCIFOR – Centro de Ciências Forenses, Portugal

ARTICLE INFO

Article history:

Keywords:

Mitochondrial DNA

Control region

Angola

Lisboa

ABSTRACT

Portugal has been considered a country of emigrants, nevertheless in the past decades the number of immigrants has grown throughout all the country. This migratory flux has contributed to a raise of heterogeneity at multiple levels. According to statistical data, at the end of 2012 the total number of Angolan immigrants in Portugal equalled about 20,000 individuals. A territorial predominance has been found for the metropolitan region of Lisboa. Angola is a country located in the Atlantic coast of Africa. The presence of Bantu people and the colonisation by Portuguese people on Angolan territory are considered to be the major modulators of the genetic patterns in Angola.

Mitochondrial DNA is known for its features that enable an approach to the study of human origin and evolution, as well to the different migration pathways of populations. This genetic marker can also contribute to ascertaining the identity of individuals in forensic cases. The main aim of this study was to determine the genetic structure of the Angolan immigrant population living in Lisboa. Therefore, a total of 173 individuals, inhabitants in Lisboa, unrelated and with Angolan ancestry were studied. Total control region of mitochondrial DNA was amplified from position 16,024 to position 576 using two pairs of primers – L15997/H016 and L16555/H639.

The majority of the identified haplotypes belong to mtDNA lineages known to be specific of the sub-Saharan region. Our results show that this immigrant population inhabitant in Lisboa presents a genetic profile that is characteristic of African populations. This study also demonstrates the genetic diversity that this immigrant population introduces in Lisboa. This does not contradict the historical data concerning colonization of Angola, since this was made mainly by male European individuals, who did not contribute with their maternal information of mtDNA.

Lisboa immigrant population from Angola can be accessed via EMPOP dataset with accession number EMPOP662.

© 2014 Published by Elsevier Ireland Ltd.

1. Introduction

Angola is a country located in the south Atlantic coast of Africa. The Khoe-san people were the first known population living in this territory. Around 3000–4000 years ago, the Bantu people, arising

from the East and Central regions of Africa, began two main migratory fluxes to the South and West of the continent. Part of this movement culminated with their fixation in Angola territory. In the 15th century, with the ascension of King D. João II to the Portuguese throne and his concern on the discovery of the sea route to Índia, the recognition of the African coast progressed rapidly. In 1482, the Portuguese navigator Diogo Cão, arrived at to the Angolan territory. Later, in 1641, Angola was occupied by Netherlands, invasion that ended in 1648, when Portugal freed the territory. Angola officially became a Portuguese colony in 1886 and it became independent in 1975 [1–3]. Thus, the arrival of Bantu

* Corresponding author at: Instituto Nacional de Medicina Legal e Ciências Forenses, Serviço de Genética e Biologia Forenses – Sul, Rua Manuel Bento de Sousa, 3, 1169-201 Lisboa, Portugal. Tel.: +351 218 811 865; fax: +351 21 8864 493.

E-mail address: amorim@dinml.mj.pt (A. Amorim).

people and the miscegenation between different sociocultural groups that occurred during the colonial time are considered the main modulators of the genetic portrait of Angolan population.

Since the beginning of the 20th century, Portugal has been known by its number of emigrant people. Nevertheless, this scenery changed in the 70s with the increasing flow of immigrants, largely related with the post-colonial and post-independence period of African countries of Portuguese language. According to statistical data from *Serviço de Estrangeiros e Fronteiras* concerning to 2012 [4], the total number of immigrant people in Portugal equalled 417,042 individuals. One of the most representative nationalities was Angolan with a total of 20,366 individuals living in Portugal; also it was been found a territorial predominance for the metropolitan region of Lisboa was also found, with 12,704 Angolan residents [4,5].

The mitochondrial DNA (mtDNA) characteristics such as maternal inheritance [6], high copy number [7], absence of recombination [8] and high mutation rate make this genetic marker useful in different fields: in medical genetics, by analysis of pathogenic mutations; in population genetics, when investigating the differences between human populations; and in forensic genetics, by comparison between one or more evidence mtDNA sequences with a reference sequence. The study of the mitochondrial genome allows ascertaining the differences between human populations in evolutionary studies; thus, it is possible to estimate the geographic origin of a population, their dynamic and the time elapsed since the most common ancestor of modern humans. The discrimination power of mtDNA is due to the polymorphic status of its hypervariable regions located in the control region – HVI, HVII and HVIII (Hypervariable region I, II and III) – with a total length of 1122 bp [9]. In the past decades, with the development of new sequencing technologies of high performance, the study of mtDNA total control region has become more common.

Although mtDNA has a lower discrimination power when compared with nuclear DNA, since the haplotype of an individual is shared within the same maternal lineage, the study of mtDNA is also commonly used in forensic genetic. In particular cases where the samples have a small amount of nuclear DNA or it is degraded into small fragments, this genetic marker reveal to be very useful [10–12].

The phenomenon of heteroplasmy – the presence of more than one type of mtDNA in one individual – has been reported in different studies. The heteroplasmy can occur in two forms: length heteroplasmy and sequence heteroplasmy. The first one is associated with insertions or deletions in homopolymeric C-stretches in HVI, HVII and HVIII regions. The second type of heteroplasmy is a result of substitutions, with presence of two different nucleotides in a single position [13–18].

The haplotype of each individual is defined by a list of polymorphic sites, comparatively to revised Cambridge Reference Sequence (rCRS) [19,20], a reference sequence, which implied that the remaining positions, not described, are equal to the rCRS. The set of polymorphisms of a given sequence of mtDNA forms a group of lineages next to each other that cluster in haplogroups, defined by the letters of the alphabet. Sequences that belong to the same haplogroup or sub-haplogroup have specific mutations, representative of a common ancestor. These haplogroups can be related with a specific geographic region or, in some cases, with an ethnic group. Thus, the study of mtDNA and its haplogroups of different populations have been used to determine the time and routes of human migration and to investigate pre historical and historical times of humanity [21,22].

An African origin of modern humans explains the increasing interest in different studies of populations with African origins [23]. Studies on maternal ancestry of several lineages based on mtDNA also allow to obtain complementary information

concerning the origin of man in Africa and its dispersion on and from this continent. The model most accepted in the scientific community concerning this matter, and confirmed by mtDNA analysis, is known as Out of Africa theory [24]. This postulates that, after an initial expansion of populations from Africa to Eurasia, the different populations did not maintain contact, and evolved independently. Subsequently, the modern man emerged and evolved in Africa, from a single population of archaic *Homo sapiens*; after this stage of evolution, a new expansion resulted in the replacement and global extinction of the population for modern human, as a result of their superior constitution [25].

The high frequency of haplogroups L shown in African populations, as well as the high level of heterogeneity of the mitochondrial DNA sequences in this continent have been reported several times, being higher than in other geographical regions [26–40].

Studies regarding mtDNA total control region of African populations have been performed [31–33,37]. However, sequences of mtDNA total control region of Angolan individuals are not published. The only information, to date, on maternal inheritance of this population includes sequences of HVI and HVII mtDNA control region [34,35,41,42], or HVI region [43]. The Angolan immigrant population living in Lisboa has been studied with STR genetic markers [44].

The aims of this study include (a) identification of the mtDNA haplotypes of the Angolan immigrant population living in Lisboa, (b) obtainment of mtDNA variability within this group of samples (c) determination of the haplogroups and (d) comparison of the present population with other African populations, hoping to understand the present and future impact of the Angolan immigrant population in Lisboa.

2. Materials and methods

2.1. Sampling and DNA extraction

Blood samples were collected from 173 Angolan immigrant inhabitants in Lisbon, non-related, of a single generation, which attended the National Institute of Legal Medicine and Forensic Science (Instituto Nacional de Medicina Legal e Ciências Forenses – INMLCF), southern branch, for kinship investigations. According to the Portuguese law that regulates INMLCF's activity, samples from routine forensic cases can be used for investigation purposes. All samples are used with special codification without any connection to judicial data related to the downer. DNA was extracted using Chelex[®] 100 method [45].

2.2. mtDNA typing

Control region of the mtDNA was amplified using two pairs of primers – L15997/H016 and L16555/H639. The amplified products were purified with ExoSap-IT (Affymetrix). Total control region sequencing was performed with BigDye[®] Terminator v.3.1 Cycle Sequence (AB) followed by purification with BigDye[®] XTerminator Purification Kit (AB). Finally, the sequenced products were detected in a sequencer Genetic Analyzer 3130 (AB) and the results were analysed by Sequencing Analysis v.5.2 and SeqScape v.3 (AB) softwares.

2.3. Sequences alignment and haplogroup classification

Obtained sequences were compared with rCRS [20] applying the nomenclature guidelines used in mtDNA analysis to identify haplotypes [46]. Haplogroups were determined based on Phylo-tree, build 16, developed by van Oven and Kayser, last updated in February 2014 [47].

2.4. Data analysis

The statistical analysis of the mtDNA total control region, calculating the sequence diversity and nucleotide diversity, as well the population diversity among other populations [31–33, 35–38,42,48], was performed using Arlequin v 3.5.1.2 software [49]. The phylogenetic representation of genetic distances was drawn with Neighbour method of Phylip v. 3.69 [50] and with Treeview v. 1.5.2 [51] softwares.

3. Results and discussion

3.1. Haplotype composition

The haplotypes acquired in this study were submitted and accepted on EMPOP database (www.empop.org) [52] accession number EMPOP662.

Along the 173 studied individuals, 145 different haplotypes were identified, 125 of which are unique. The other 20 haplotypes were present in more than one individual. Thus, 14 sequences shared their haplotypes within groups of 2 samples, 4 sequences within groups of 3 samples and 2 sequences within groups of 4 samples. The two most frequent haplotypes are 16185T 16223T 16327T 16519C 73G 150T 189G 200G 263G 309.1C 315.1C (ANG078, ANG151, ANG166 and ANG171) and 16129A 16148T 16168T 16172C 16187T 16188G 16189C 16223T 16230G 16278T 16293G 16311C 16320T 93G 95C 185A 189G 236C 247A 263G 309.1C 315.1C 523DEL 524DEL (ANG012, ANG077, ANG106 and ANG152) present in four samples each.

The transition of a cytosine (C) for a thymine (T) in position 16,223 of HVI region, it is present in 153 samples. In 142 sequences was observed a transition in position 73 of HVII region was observed, where an adenine (A) is replaced by a guanine (G). In HVII region, position 152, a substitution of a thymine (T) by a cytosine (C) is observed in 103 sequences. Also, in position 16,278 of HVI, a cytosine (C) is replaced by a thymine (T) in 92 sequences. The most frequent polymorphisms present in this sampling are also found in other African populations [31–33].

In the 173 studied individuals one point heteroplasmy was identified at position 16,183 of HVI region in sample ANG163. Here, in some mtDNA molecules of the individual, the nucleotide present is an Adenine, while in others is a Cytosine, being the position classified as 16183M. Regarding length heteroplasmy, associated with homopolymeric C-stretches in HVI, HVII and HVIII regions, there are a total of 55 heteroplasmic samples. In 25 samples it is present only in HVI region, in 19 it is present only in HVII region and in 1 sample it is present only in HVIII region. However, in 8 samples the length heteroplasmy is present in both HVI and HVII, and in 2 samples it is present in HVI and HVIII regions. The remaining 118 samples are homoplasmic.

3.2. Haplogroup composition

Macrohaplogrup L is the most common haplogroup in our sample, since there are 151 samples that fit into L haplogroup (about 87%). Haplogroups H and T are present in 5 samples each. Haplogroup R0 is present in 4 samples, haplogroups K and U appear in 3 samples each and the haplogroups J and M are present in one sample each.

The L lineages are known to be characteristic from Sub-Saharan regions of Africa, and it is in this continent that it is found the higher genetic diversity [27,39,40]. Within haplogroup L0 there are four sub-haplogroups - L0a, L0d, L0f and L0k. In this study the samples presented haplotypes belonging to sub-haplogroups L0d (1 sequence) and L0a (23 sequences). The sub-haplogroup L0d is known for being restricted to Khoe-san people in South Africa

and to Tanzania and Angolan populations [39,40,53]; on the other hand, the L0a lineages are widely spread through Africa (eastern, central and southern areas) [39,40]. The Khoe-san L0k was not found in our samples. Regarding to haplogroup L1, five samples were included in L1b, characteristic of western-central African regions [28,40]. The L1c, which is concentrated in Central and West of Africa includes 32 of our samples. The haplogroups L3 and L2 comprise about 70% of the Sub-Saharan maternal variation [40]. Three of the five subclades of haplogroup L2 (L2a, L2b, L2c) were present in 32 of our studied individuals; these are known for existing in the West-East region of Africa. The lineage L3 is spread through Africa [39,40]; a total of 56 of our studied individuals were within this haplogroup, included in L3a, L3b, L3d and L3f subclades. The L4 lineages, despite having low frequencies, are mostly found in the East-North of Africa; in this case, 2 of our samples belong to this haplogroup [39,40].

Although present in a small range, some of the haplogroups identified in our samples, do not belong to L lineages and are not characteristic from the African population, even though they can be found in some areas of this continent. The European populations derive from haplogroup N, which branches into haplogroups H, I, J, K, T, U, V, W and X. On the other hand, the Asiatic lineages belong both to haplogroups N and M.

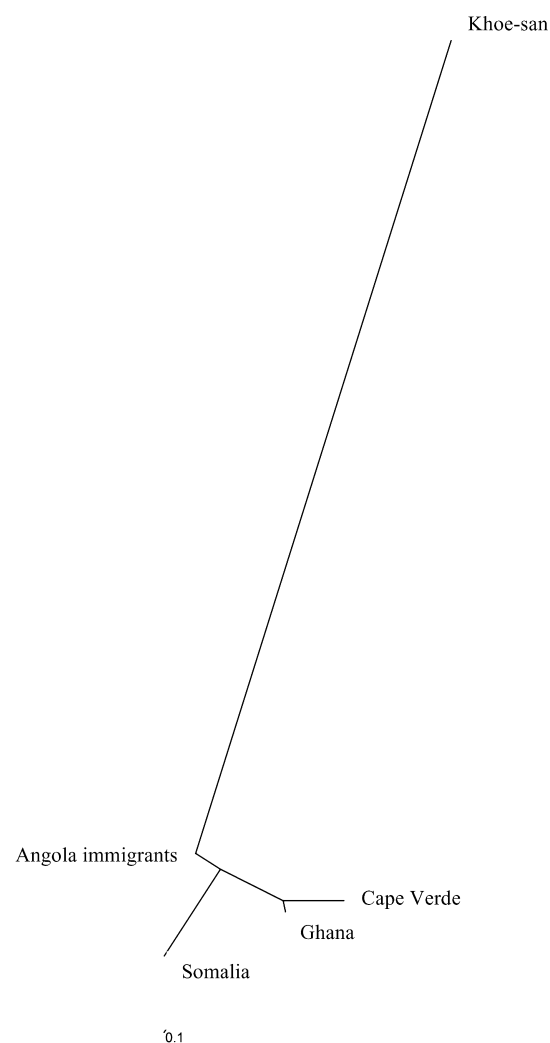


Fig. 1. Phylogenetic representation of the genetic distances for African populations regarding mtDNA total control region.

3.3. Genetic diversity

Statistical analyses were made taking into consideration the total control region of sequences. According to calculations obtained from Arlequin software v. 3.5.2.1 [49], for this population the value of the sequence diversity is 0.9964 ± 0.0012 and the nucleotide diversity is 0.016028 ± 0.007910 . These values were also calculated for HVI and HVII regions, in order to be able to compare our results with the published studies concerning only to HVI and HVII mtDNA control region. Therefore the sequence diversity for HVI and HVII regions of this population is 0.9959 ± 0.0012 and the nucleotide

diversity is 0.025058 ± 0.012429 . The values of the sequence diversity are higher when the total control region is studied, in comparison with the study of the hypervariable regions HVI and HVII.

3.4. Inter population analysis

In Fig. 1 we present the phylogenetic tree based on molecular distances between populations, calculated with Arlequin v. 3.5.2.1 software [49], for the mtDNA total control region. Our Lisboa immigrant population from Angola was compared with

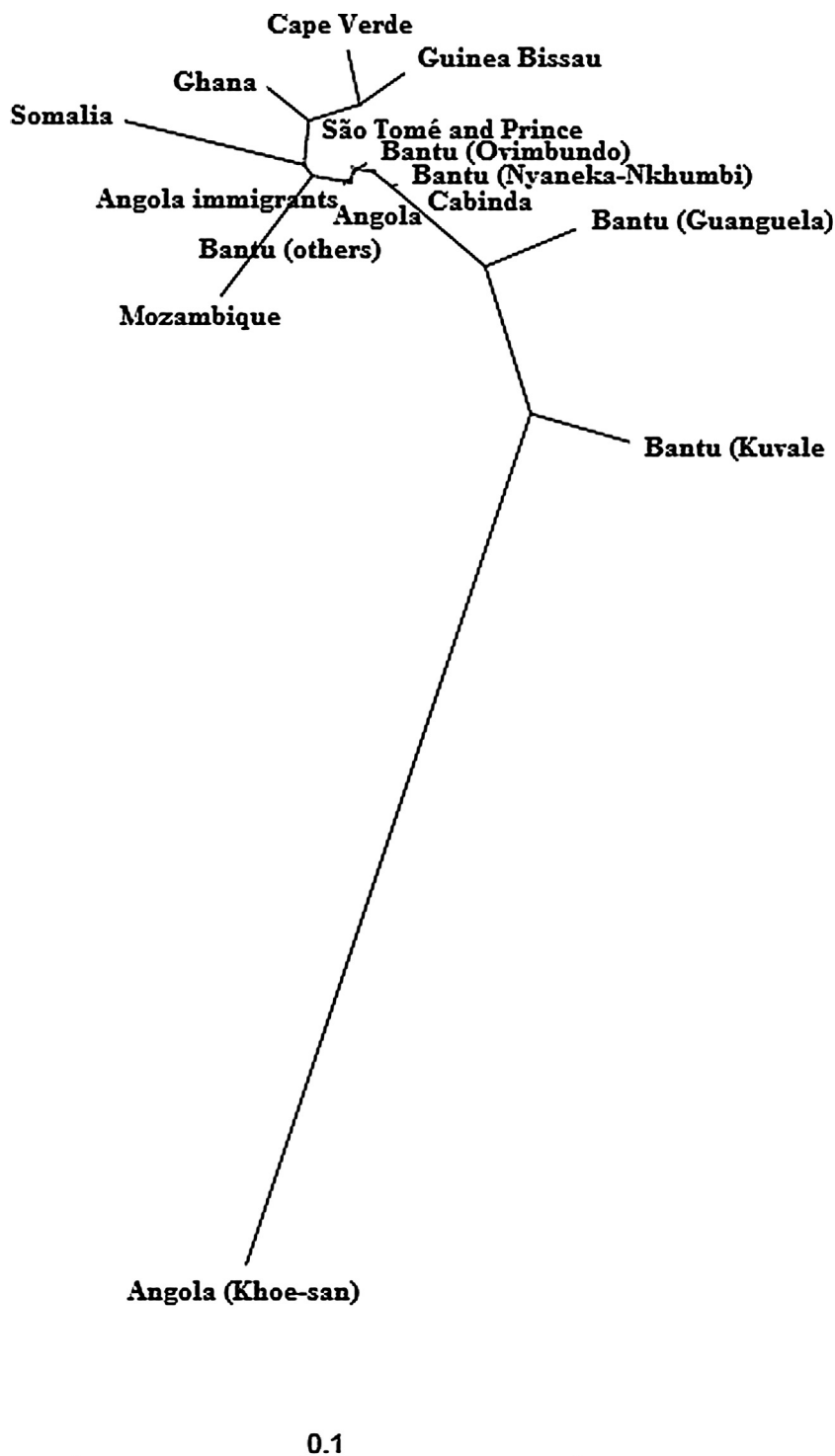


Fig. 2. Phylogenetic representation of the genetic distances for African populations regarding HVI and HVII mtDNA control region.

Table 1

Populations used in the phylogenetic study of mtDNA total control region.

Population	Sample size	References
Present study	173	–
Khoe-san	129	Freundt (2012)
Ghana	192	Freundt (2012)
Somalia	190	Mikkelsen (2012)
Cape Verde	103	Morais (2013)

Table 2Analysis of molecular variance (AMOVA): pairwise difference genetic distance (FST/FST ρ values) between the studied population and other populations from the literature, regarding mtDNA total control region.

Populations	Sample size	Fst	ρ
Present study	173	0.00000	–
Khoe-san	129	0.24677	0.00000 \pm 0.0000
Cape Verde	103	0.04408	0.00000 \pm 0.0000
Somalia	190	0.03588	0.00000 \pm 0.0000
Ghana	192	0.02746	0.00000 \pm 0.0000

Table 3

Populations used in the phylogenetic study of mtDNA HVI and HVII regions.

Population	Sample size	References
Present study	173	–
Angola (Khoe-san)	129	Freundt (2012)
Ghana	192	Freundt (2012)
Somalia	190	Mikkelsen (2012)
Cape Verde	103	Morais (2013)
Guinea Bissau	88	Carvalho (2011)
Mozambique	109	Pereira (2001)
São Tomé and Prince	100	Trovoada (2004)
Angola	59	Carvalho (2006)
Bantu (others)	45	Coelho (2009)
Bantu (Kuvale)	54	Coelho (2009)
Bantu (Guanguela)	21	Coelho (2009)
Bantu (Nyaneka-Nkhumbi)	153	Coelho (2009)
Bantu (Ovimbundu)	92	Coelho (2009)
Cabinda	110	Beleza (2005)

populations from Ghana [32], Somalia [31], Cape Verde [33] and from indigenous Angolan Khoe-san [37] (Table 1). Although the immigrant Angolans living in Lisboa show a great distance from Khoe-san population (Fst = 0.24677; ρ = 0.00000), it is smaller than the distance between the latter one and the other populations analysed in the study. Still, the Angolan population living in Lisboa is closer to Ghana population (Fst = 0.02746; ρ = 0.00000), followed by Somalia (Fst = 0.03588; ρ = 0.00000) and Cape Verde (Fst = 0.04408; ρ = 0.00000) (Table 2). The lack of Khoe-san haplogroups contribution into southwest regions of Africa has already been described by other authors [26,36,54].

For comparison purposes, and as there are few studies regarding African population for mtDNA total control region, the sequences of the studied population were reduced to HVI and HVII regions, to allow comparison with other populations studied for HVI and HVII regions. In this case, the population of Angolan immigrants living in Lisboa was compared with populations from Guinea Bissau [29], Cape Verde [33], Somalia [31], Ghana [32], Mozambique [38], S. Tome and Principe [48], Angola [42], indigenous Khoe-san [37], Bantu ethnic groups from Angola [35,36] and Cabinda [34], all of them with sequenced HVI and HVII regions (Table 3) (Fig. 2). These results culminate in a short distance between the population of Angolan immigrants living in Lisboa and Angolan, people living in Angola as well Bantu ethnic groups from Angola, giving information that the first one maintains the genetic information of its origin, although, nowadays, inhabiting a different geographic region.

4. Conclusions

The study of the entire control region for the Angolan immigrant population living in Lisboa, obtained for the first time, increases the discrimination power of the sequences and allows a better determination of haplogroups. Due to the high frequency of unique haplotypes present in our samples, the population has a great genetic variability.

It is possible to verify the genetic similarity with African populations. From the analysis restricted to HVI and HVII regions it was determined that the Angolan immigrants living in Lisboa have a small distance to the native Angolan populations. These results were expected since, according to historical data, the colonisation of this country by the European population it was made mainly by male individuals, which would lead to a low contribution to European mtDNA sequences of possible miscegenation during colonial times. On the other hand, the presence of these immigrants in Portugal is relatively recent, which eventually results in a lower contribution of the haplotypes in our new generations; but in a closer future, these African haplotypes will, eventually, be a part of the overall Lisboa population haplotypes. On the other hand, the great distance of our studied population from Khoe-san population demonstrates the high replacement rate of Khoe-san people by Bantu populations in Angolan territory.

Acknowledgements

The authors would like to thank Walther Parson and Alexander Röck (EMPOP, University of Innsbruck, Austria).

References

- [1] J. Redinha, *Etnias e culturas de Angola*, 2009.
- [2] J. De Raeymaeker, *A descoberta de Angola*, 1st ed., Oficina do Livro – Sociedade Editorial Lda, Lisboa, 2012.
- [3] *Iniciação à história da Angola*, República Popular de Angola, Ministério da Educação, 1992.
- [4] P. Dias, R. Machado, J. Estrela, A.R. Bento, *Relatório de imigração fronteiras e asilo* 2012, Oeiras, 2012.
- [5] PORDATA: Base de Dados Portugal Contemporâneo, 2014.
- [6] R.E. Giles, H. Blanc, H.M. Cann, D.C. Wallace, Maternal inheritance of human mitochondrial DNA, *Proc. Natl. Acad. Sci. U. S. A.* 77 (1980) 6715–6719.
- [7] H.H. Cells, D. Bogenhagen, D.A. Clayton, The number of mitochondrial deoxyribonucleic acid genomes in mouse and human HeLa cells, *J. Biol. Chem.* 249 (1974) 7991–7995.
- [8] J. Williams, O. Ballard, M.C. Whitlock, The incomplete natural history of mitochondria, *Mol. Ecol.* 13 (2004) 729–744.
- [9] J.M. Butler, *Forensic DNA Typing*, 2005.
- [10] S. Lutz, H.J. Weissner, J. Heizmann, S. Pollak, mtDNA as a tool for identification of human remains. Identification using mtDNA, *Int. J. Legal Med.* 109 (1996) 205–209.
- [11] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, *Int. J. Legal Med.* 108 (1995) 68–74.
- [12] W. Parson, T.J. Parsons, R. Scheithauer, M.M. Holland, Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case, *Int. J. Legal Med.* 111 (1998) 124–132.
- [13] D. Comas, S. Paabo, J. Bertranpetit, Heteroplasmy in the control region of human mitochondrial DNA, *Genome Res.* 5 (1995) 89–90.
- [14] K.A. Roberts, C. Calloway, Characterization of mitochondrial DNA sequence heteroplasmy in blood tissue and hair as a function of hair morphology, *J. Forensic Sci.* 56 (2011) 46–60.
- [15] A. Alonso, A. Salas, C. Albarrán, E. Arroyo, A. Castro, M. Cresillo, et al., Results of the 1999–2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the CEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor, *Forensic Sci. Int.* 125 (2002) 1–7.
- [16] A. Brandstätter, W. Parson, Mitochondrial DNA heteroplasmy or artefacts – a matter of the amplification strategy? *Int. J. Legal Med.* 117 (2003) 180–184.
- [17] N. Howell, C.B. Smejkal, Persistent heteroplasmy of a mutation in the human mtDNA control region: hypermutation as an apparent consequence of simple-repeat expansion/contraction, *Am. J. Hum. Genet.* 66 (2000) 1589–1598.
- [18] M.R. Wilson, D. Polanskey, J. Replogle, J.A. DiZinno, B. Budowle, A family exhibiting heteroplasmy in the human mitochondrial DNA control region reveals both somatic mosaicism and pronounced segregation of mitotypes, *Hum. Genet.* 100 (1997) 167–171.

- [19] S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, et al., Sequence and organization of the human mitochondrial genome, *Nature* 290 (1981) 457–465.
- [20] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nat. Genet.* 23 (1999) 147.
- [21] H.J. Bandelt, M. van Oven, A. Salas, Haplogrouping mitochondrial DNA sequences in legal medicine/forensic genetics, *Int. J. Legal Med.* 126 (2012) 901–916.
- [22] B. Pakendorf, M. Stoneking, Mitochondrial DNA and human evolution, *Annu. Rev. Genomics Hum. Genet.* 6 (2005) 165–183.
- [23] C.M. Schlebusch, P. Skoglund, P. Sjödin, L.M. Gattepaille, D. Hernandez, F. Jay, et al., Genomic variation in seven Khoe-San groups reveals adaptation and complex African history, *Science* 338 (2012) 374–379.
- [24] N. Maca-Meyer, A.M. González, J.M. Larruga, C. Flores, V.M. Cabrera, Major genomic mitochondrial lineages delineate early human expansions, *BMC Genet.* 2 (2001) 13.
- [25] I. Tattersall, Human origins: out of Africa, *Proc. Natl. Acad. Sci. U. S. A.* 106 (2009) 16018–16021.
- [26] A. Salas, M. Richards, M.-V. Lareu, R. Scozzari, A. Coppa, A. Torroni, et al., The African diaspora: mitochondrial DNA and the Atlantic slave trade, *Am. J. Hum. Genet.* 74 (2004) 454–465.
- [27] Y.S. Chen, A. Torroni, L. Excoffier, A.S. Santachiara-Benerecetti, D.C. Wallace, Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups, *Am. J. Hum. Genet.* 57 (1995) 133–149.
- [28] E. Watson, K. Bauer, R. Aman, G. Weiss, A. von Haeseler, S. Pääbo, mtDNA sequence diversity in Africa, *Am. J. Hum. Genet.* 59 (1996) 437–444.
- [29] M. Carvalho, P. Brito, A.M. Bento, V. Gomes, H. Antunes, H.A. Costa, et al., Paternal and maternal lineages in Guinea-Bissau population, *Forensic Sci. Int. Genet.* 5 (2011) 114–116.
- [30] M. Carvalho, V. Bogas, F. Balsa, H.A. Costa, V. Lopes, M.M. Melo, et al., Mitochondrial control region data of 3 ethnic groups from angola, *Forensic Sci. Int. Genet.* 3 (2011) e125–e126.
- [31] M. Mikkelsen, L. Fendt, A.W. Röck, B. Zimmermann, E. Rockenbauer, A.J. Hansen, et al., Forensic and phylogeographic characterisation of mtDNA lineages from Somalia, *Int. J. Legal Med.* 126 (2012) 573–579.
- [32] L. Fendt, A. Röck, B. Zimmermann, M. Bodner, T. Thyse, F. Tschentscher, et al., Mtdna diversity of Ghana: a forensic and phylogeographic view, *Forensic Sci. Int. Genet.* 6 (2012) 244–249.
- [33] P. Morais, Estudo do DNA mitocondrial de um grupo de imigrantes oriundos de Cabo Verde residentes em Lisboa Mestrado em Biologia Molecular em Saúde, 2013.
- [34] S. Belez, L. Gusmão, A. Amorim, A. Carracedo, A. Salas, The genetic legacy of western Bantu migrations, *Hum. Genet.* 117 (2005) 366–375.
- [35] M. Coelho, F. Sequeira, D. Luiselli, S. Belez, J. Rocha, On the edge of Bantu expansions: mtDNA, Y chromosome and lactase persistence genetic variation in southwestern Angola, *BMC Evol. Biol.* 9 (2009) 80, <http://dx.doi.org/10.1186/1471-2148-9-80>.
- [36] S. Plaza, A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, A. Carracedo, et al., Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, *Hum. Genet.* 115 (2004) 439–447.
- [37] L. Fendt, G. Huber, A.W. Röck, B. Zimmermann, M. Bodner, R. Delpont, et al., Mitochondrial DNA control region data from indigenous Angolan Khoe-San lineages, *Forensic Sci. Int. Genet.* 6 (2012) 662–663.
- [38] L. Pereira, V. Macaulay, A. Torroni, R. Scozzari, Z. Pavia, V. Ferrata, Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansion and the slave trade, *Ann. Hum. Genet.* 65 (2001) 439–458.
- [39] A. Salas, M. Richards, T. De la Fe, M.-V. Lareu, B. Sobrino, P. Sánchez-Diz, et al., The making of the African mtDNA landscape, *Am. J. Hum. Genet.* 71 (2002) 1082–1111.
- [40] A. Rosa, A. Brehem, African human mtDNA phylogeography at-a-glance, *J. Anthropol. Sci.* 89 (2011) 25–58.
- [41] S. Plaza, A. Salas, F. Calafell, F. Corte-Real, J. Bertranpetit, A. Carracedo, et al., Insights into the western Bantu dispersal: mtDNA lineage analysis in Angola, *Hum. Genet.* (2004) 439–447.
- [42] M. Carvalho, F. Balsa, V. Lopes, M.J. Anjos, L. Andrade, A. Serra, et al., Mitochondrial DNA analysis in a population from Angola, in: XXTH Congress of International Academy of Legal Medicine, 23–26 August, Budapest, Hungary, 2006.
- [43] S. Tofanelli, P. Garagnani, C. Bini, X. Fosella, S. Pelotti, G. Paoli, et al., mtDNA variability in two Bantu-speaking populations (Shona and Hutu) from Eastern Africa: implications for peopling and migration patterns in Sub-Saharan Africa, *Am. J. Phys. Anthropol.* 311 (2009) 302–311.
- [44] R.M. dos Santos, C.V. da Silva, H.A. Costa, P.F. Gomes, S. Sanches, R. Espinheira, et al., Genetic portrait of an immigrant population from Angola living in Lisboa, *Forensic Sci. Int. Genet.* 6 (6) (2012), e170–3.
- [45] P.S. Walsh, D.A. Metzger, R. Higuchi, Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material, *Biotechniques* 10 (1991) 506–513.
- [46] A. Carracedo, W. Bar, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, et al., DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing, *Forensic Sci. Int.* 110 (2000) 17–85.
- [47] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) E386–E394.
- [48] M.J. Trovada, L. Pereira, L. Gusmão, A. Abade, A. Amorim, M.J. Prata, Insights from pattern of mtDNA variation into the genetic history of São Tomé e Príncipe, *Int. Congr. Ser.* 1261 (2004) 377–379.
- [49] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0); an integrated software package for population genetics data analysis, *Evol. Bioinform. Online* 1 (2005) 47–50.
- [50] J. Felsenstein, PHYLIP: phylogeny interface package (version 3.2), *Cladistics* 5 (1989) 164–166.
- [51] R.D.M. Page, Treeview: an application to display phylogenetic trees on personal computers, *Comput. Appl. Biosci.* 12 (1996) 357–358.
- [52] W. Parson, A. Dür, EMPPOP – a forensic mtDNA database, *Forensic Sci. Int. Genet.* 1 (2007) 88–92.
- [53] D.M. Behar, R. Villems, H. Soodyall, J. Blue-smith, L. Pereira, E. Metspalu, et al., The dawn of human matrilineal diversity, *Am. J. Hum. Genet.* 82 (2008) 1130–1140.
- [54] J. Alves-Silva, M. da Silva Santos, P.E. Guimarães, A.C. Ferreira, H.J. Bandelt, S.D. Pena, et al., The ancestry of Brazilian mtDNA lineages, *Am. J. Hum. Genet.* 67 (2000) 444–461.